Statistical Signal and Image Processing Techniques in Corneal Modeling

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Weaam F. H. Alkhaldi, M. Sc.
geboren am 10.06.1974 in Alexandria, Ägypten

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Dr. D. Robert Iskander
Korreferent:	Prof. Dr. Henryk Kasprzak
Tag der Einreichung:	22.04.2010
Tag der mündlichen Prüfung:	18.06.2010

D 17
Darmstädter Dissertation
Darmstadt, 2010

Acknowledgments

This thesis was a result of years of research that included turbulent times, but also times of happiness and success. I am greatly indebted to many people, who have provided endless support and help.

First, I wish to express my sincerest gratitude to the German Academic Exchange Service (DAAD). They have offered me the great opportunity to come to Germany and start a new step in my life. Throughout the years, they were always providing support and keeping up the spirit. Particularly, I would like to thank Dr. Baumgarten, Mrs. Muhareb from the DAAD office in Jerusalem, Palestine, and Mrs. Hanzlik, Mrs. Bönninghausen, Mrs. Gerecke, and Mrs. Barth from the DAAD headquarter in Bonn, Germany.

Coming to the work's environment, I would like to deeply thank my first supervisor, Prof. Abdelhak M. Zoubir, for giving me the great chance to work in the Signal Processing Group besides invaluable support, significant guidance and encouragement. I highly appreciate the warm-hearted brotherhood's feelings he has shown and his to-the-point advices. On the other side of the globe, in Australia, there has been Dr. D. Robert Iskander, who has made the course of this work really fruitful and enjoyable. I am very grateful for his unlimited assistance and great discussions, both technically and personally. Also, many thanks go to Prof. Henryk Kasprzak for co-supervising this thesis and the instructive comments and suggestions he has provided.

Having in mind the magnificent time I have spent in the Signal Processing Group, I would like to sincerely thank ALL my great colleagues, both former and current ones, and also the visitors, for their friendly feelings and the productive discussions we had.

Finally, I would like to dedicate this work to my family. For my father (may God bless his soul) and my mother (God bless her), I would say that without your infinite support and love from childhood to now, this achievement would not have come true at all. Also, many thanks go for my brother and sister for their support and kind feelings. Adjacently, I wish to express my love and gratitude to my wife and my two sons for their patience, support, and understanding.

Darmstadt, July 2010

Weaam Alkhaldi

Kurzfassung

In dieser Dissertation betrachten wir zwei mit einander verknüpfte Probleme: Die Verbesserung von videokeratoskopischen Bildern zur genaueren Schätzung der Hornhauttopographie, sowie die Schätzung der Modellordnung bei der Modellierung dieser mittels orthogonaler Zernike-Polynome.

Die Schätzung der Hornhauttopographie mit einem Videokeratoskop beruht auf dem Placido-Scheiben-Prinzip. Vorraussetzungen für akkurate Schätzungen sind eine hohe Qualität des präokulären Tränenfilms und ein hinreichend grosser Lidspalt zur Vermeidung von Augenlidreflektionen. Diese Bedingungen sind in der Praxis jedoch nicht immer erfüllt, denn einerseits leiden einige Patienten an einer Austrocknung des äuseren Auges, und andererseits haben, vor allem Patienten asiatischer Abstammung, häufig einen kleinen Lidspalt. Eine schlechte Qualität des Tränenfilms und Reflektionen der Wimpern, führen zu Datenverlusten in bestimmten Regionen und einem verminderten Erfassungsbereich durch die Placido-Scheibe. Dies resultiert in einer schlechteren Schätzung der Hornhauttopographie für diese Patienten. Unser Ziel ist, durch eine Vergrösserung des Standarderfassungsbereichs von Placido-Scheiben-Videokeratoskopen, auch in den oben genannten Fällen, eine zuverlässige Schätzung der Hornhauttopgraphie zu ermöglichen. Dies erreichen wir durch die Einbindung von linearer adaptiver Filterung (basierend auf der Theorie der Wiener-Filter), sowie nichtlinearer Filterung (basierend auf morphologischen Operationen) von Einzelbildern, in den Topographieschätzungs-Algorithmus des Instruments. Die Ergebnisse der Messungen von Referenzoberflächen und echten Hornhäuten belegen, dass die Einbeziehung der vorgeschlagenen Techniken zu einer verbesserten Schätzung der Hornhauttopographie, und in vielen Fällen zu einer deutlichen Vergrösserung des Erfassungsbereiches, und damit, zu einem verbesserten Videokeratoskop für die Fachärzte, führt.

Die mathematische Modellierung der Höheninformation der Hornhauttopographie erfolgt in der Regel durch eine Reihe von orthogonalen Zernike-Polynomen. Wir betrachten hier die Schätzung der Anzahl der notwendigen Zernike-Polynome als Modell Auswahl-Problem, welches in der linearen Regression formalisiert ist. Klassische informationstheoretische Kriterien zur Modell Auswahl neigen dazu die Komplexität der Hornhautoberfläche durch einen zu schwachen Strafterm zu überschätzen, während Bootstrap-Techniken diese unterschätzen, oder einen hohen Verarbeitungsaufwand benötigen. Wir schlagen vor, das Effiziente Detektionskriterium (EDC), das die gleiche allgemeine Form wie die informationstheoretischen Kriterien hat, als alternative Methode zur Modellordnungs-Schätzung zu verwenden. Wir zeigen zunächst durch Simulationen, dass das EDC einer grossen Anzahl von informationstheoretischen und

resampling-basierten Kriterien überlegen ist. Wir veranschaulichen dann, anhand von echten Hornhautdaten, dass unsere mit dem EDC erzielten Ergebnisse in enger Übereinstimmung mit klinischen Erwartungen sind, und somit das EDC ein Mittel für die Unterscheidung zwischen normalen, astigmatischen und keratokonischen Hornhautoberflächen dartellt.

Die beiden in dieser Dissertation behandelten Probleme sind dadurch miteinander verknüpft, dass die korrekte Modellierung der Hornhauttopographie, mit welcher Funktion auch immer, eine präzise Schätzung dieser vorraussetzt, welche wiederum auf einer ausreichenden Qualität der videokeratoskopischen Bilder basiert.

Abstract

In this thesis, we consider two interrelated problems, which are the enhancement of videokeratoscopic images for more accurate corneal topography estimation and model-order selection of corneal surfaces when expanded using orthogonal Zernike polynomials.

Corneal topography estimation that is based on the Placido disk principle relies on good quality of pre-corneal tear film and sufficiently wide eyelid (palpebral) aperture to avoid reflections from eyelashes. However, in practice, these conditions are not always fulfilled resulting in missing regions, smaller corneal coverage, and subsequently poorer estimates of corneal topography. Our aim is to enhance the standard operating range of a Placido disk videokeratoscope to obtain reliable corneal topography estimates in patients with poor tear film quality, such as encountered in those diagnosed with dry eye, and with narrower palpebral apertures as in the case of Asian subjects. This is achieved by incorporating in the instrument's own topography estimation algorithm an image processing technique that comprises of linear adaptive filter (based on Wiener filtering theory) and non-linear filter (based on morphological operations). The experimental results from measurements of test surfaces and real corneas show that the incorporation of the proposed technique results in better estimates of corneal topography and, in many cases, to a significant increase in the estimated coverage area making such an enhanced videokeratoscope a better tool for clinicians.

On the other hand, corneal height-data are typically modeled using a set of orthogonal Zernike polynomials. We address the estimation of the number of Zernike polynomials, which is formalized as a model-order selection problem in linear regression. Classical information theoretic criteria tend to overestimate the corneal surface due to the weakness of their penalty functions, while bootstrap-based techniques tend to underestimate the surface or require extensive processing. We propose to use the Efficient Detection Criterion (EDC), which has the same general form of information theoretic-based criteria, as an alternative to estimating the optimal number of Zernike polynomials. We first show, via simulations, that the EDC outperforms a large number of information theoretic criteria and resampling-based techniques. We then illustrate that using the EDC for real corneas results in models that are in closer agreement with clinical expectations and provides means for distinguishing normal corneal surfaces from astigmatic and keratoconic surfaces.

The two problems are interrelated in the sense that appropriate modeling for corneal surfaces, regardless of the used functions, requires accurate corneal topography that can be efficiently estimated provided that the videokeratoscopic image is not degraded.

Publications

Internationally Refereed Journal Articles

- W. Alkhaldi, D. R. Iskander, and A. M. Zoubir, “Model-order selection in Zernike polynomial expansion of corneal surfaces using the efficient detection criterion,” *IEEE Transactions on Biomedical Engineering*, 2010, in press.
- W. Alkhaldi, D. R. Iskander, A. M. Zoubir, and M. J. Collins, “Enhancing the standard operating range of a placido disk videokeratoscope for corneal surface estimation,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 800–809, March 2009.

Internationally Refereed Conference Papers

- W. Alkhaldi, D. R. Iskander, and A. M. Zoubir, “Improving the performance of model-order selection criteria by partial-model selection search,” in *The Proceedings of The 35th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Texas, USA, March 2010, pp. 4130–4133.
- W. Alkhaldi, D. R. Iskander, and A. M. Zoubir, “The applicability of biased estimation in model and model order selection,” in *The Proceedings of The 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009, pp. 3461–3464.
- D. R. Iskander, W. Alkhaldi, and A. M. Zoubir, “On the computer intensive methods in model selection,” in *The Proceedings of The 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, March 2008, pp. 3461–3464.
- D. R. Iskander and W. Alkhaldi, “The “hook and loop” resampling plane,” in *The Proceedings of The 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Virgin Islands, USA, December 2007, pp. 65–68.

Contents

1 **Introduction** **1**
- 1.1 Aims and Research Objectives . . . 1
- 1.2 Summary of the Contributions . . . 3
- 1.3 Overview of the Dissertation . . . 4

2 **Reconstruction and Modeling of Corneal Surfaces** **5**
- 2.1 Introduction and Motivation . . . 5
- 2.2 The Vision Process . . . 5
- 2.3 The Cornea . . . 7
- 2.4 Basics of Corneal Topography . . . 10
- 2.5 Placido Disk Based Videokeratoscopy . . . 12
  - 2.5.1 Basic Principle . . . 13
  - 2.5.2 Applications . . . 14
- 2.6 Corneal Modeling . . . 15
  - 2.6.1 Zernike Polynomials . . . 16
  - 2.6.2 Applications in Model-Order Selection . . . 17
- 2.7 Summary . . . 18

3 **Enhancement of Videokeratoscopic Images** **21**
- 3.1 Introduction and Motivation . . . 21
- 3.2 Methodologies . . . 24
  - 3.2.1 Centroid Detection . . . 26
  - 3.2.2 Polar Transformation . . . 30
  - 3.2.3 Adaptive Spatial Filtering . . . 34
  - 3.2.4 Morphological Closing . . . 37
  - 3.2.5 Cartesian Transformation . . . 39
- 3.3 Performance Evaluation . . . 40
- 3.4 Experimental Results with Discussion . . . 42
- 3.5 Summary . . . 55

4 **Model-Order Selection for Corneal Surfaces** **61**
- 4.1 Introduction and Motivation . . . 61
- 4.2 Model-Order Selection . . . 63
  - 4.2.1 Information Theoretic Criteria . . . 64
  - 4.2.2 Resampling-Based Criteria . . . 67
- 4.3 Simulation Results . . . 68
- 4.4 Experimental Results . . . 75

4.4.1 Corneal Discrimination 80
4.4.2 Population Match 84
4.5 Summary 85

5 Conclusions and Future Work 89
5.1 Summary and Conclusions 90
5.1.1 Enhancement of Videokeratoscopic Images for Accurate Corneal Topography Estimation 90
5.1.2 Model and Model-Order Selection 91
5.2 Future Work 93
5.2.1 Enhancement of Videokeratoscopic Images for Accurate Corneal Topography Estimation 93
5.2.2 Model and Model-Order Selection 94

Appendix A: Model and Model-Order Selection 97
A.1 Introduction and Motivation 97
A.2 The Model 99
A.3 Information-Based Model Selection Criteria 100
A.3.1 Akaike's Information Criterion 100
A.3.2 Corrected Akaike's Information Criterion 101
A.3.3 Minimum Description Length 101
A.3.4 Normalized Maximum Likelihood 102
A.3.5 The Hannan-Quinn Criterion 103
A.3.6 Conditional Model-Order Estimator 104
A.3.7 Mallows' $\mathrm{C_p}$ Criterion 105
A.3.8 Efficient Detection Criterion 106
A.4 Resampling-Based Model Selection Criteria 107
A.4.1 Bootstrap Based Model Selection 107
A.4.2 Hook-and-Loop Based Model Selection 108
A.5 Biased Estimation in Linear Models 110
A.6 Model-Order Selection by Partial-Model Selection Search 111
A.7 Simulation Results with Discussion 112
A.7.1 Biased versus Unbiased Estimation 112
A.7.1.1 Model Selection 113
A.7.1.2 Model-Order Selection: Example 1 114
A.7.1.3 Model-Order Selection: Example 2 115
A.7.2 Partial- versus Full-Model Order Selection Searching Methods . 116
A.7.2.1 Example 1 116
A.7.2.2 Example 2 117
A.7.2.3 Example 3 117

A.7.2.4 Example 4 118
A.8 Summary 119

List of Acronyms 121

List of Symbols 123

Bibliography 127

Lebenslauf 137

Chapter 1

Introduction

1.1 Aims and Research Objectives

The cornea is a transparent surface of the anterior part of the human eye, which has two major roles in vision. First, it is transparent to enable light to be transmitted to the retina. Second, its anterior surface is the primary refractive component of the human eye. With its refractive power, the cornea contributes to about 70% [1,2] of the total optical power of the eye. Therefore, it plays an important role in the refraction and the focussing of the light on the retina. Subtle variations in the corneal shape significantly affect the visual acuity.

The optical quality of the cornea is a critical factor in the proper visual function of the human eye and can be determined using corneal topography [3], which is a method of corneal shape examination assisted by computer analysis. Instruments that measure the corneal topography are called corneal topographers (e.g., videokeratoscopes). Most of the videokeratoscopes used in clinical practice today are based on the Placido disk technique. They project a series of illuminated rings onto the corneal surface, which are reflected off the cornea and captured by a video camera. The images of the reflected rings are analyzed by a computer, and different topographical maps of the cornea are generated. These maps are capable of revealing any distortions and/or irregularities of the corneal surface provided that the captured images are of a high quality. However, in some non-standard image acquisition scenarios, where the patients are not able to open their eyes wide and/or they suffer from dry-eye syndrome, the captured images are of poor quality and not reliable. Having in mind that the quality of reflected images has a significant influence on the subsequent estimation of the corneal topography, the question of interest is whether applying image processing techniques to enhance the quality of the videokeratoscopic images would yield more accurate estimation of the corneal topography. Most important is avoiding *over-enhancement* of videokeratoscopic images and its subsequent fake/inaccurate corneal topography estimation.

Zernike polynomials expansion is traditionally used as a fitting routine to model corneal topography data [4]. The coefficients of Zernike polynomials expansion have optical interpretation in the form of the basic aberrations as well as higher order aberrations. Zernike polynomials have the ability to reconstruct corneal surfaces provided that they are nonseverely aberrated surfaces. However, small irregularities cannot be captured by Zernike polynomials [5]. Therefore, how many Zernike terms one should fit to the corneal surface is an important issue (i.e., model-order selection problem). There have been many research efforts to solve this problem [4, 6–9], from which bootstrap based model-order selection [10–13] was recommended. However, efficient and computationally simple methods for model-order selection of corneal surfaces are still missed.

Our aim in this thesis is two folds. First, enhancing the videokeratoscopic images obtained by a commercial videokeratoscope. It is noteworthy to mention that the algorithms used in commercial instruments are proprietary and, to the best of our knowledge, there are only few attempts that addressed this enhancement issue in the literature [14, 15]. We suggest that this enhancement can be mainly done by applying a combined filtering technique (i.e., based on the adaptive pixel-wise Wiener filter and a non-linear (morphological) filter). The objective is to recover the missing data that can be faced during image acquisition in non-standard scenarios (e.g., reflections from eyelashes, tear-film break-ups, and existence of mucus on the corneal surface) while at the same time to keep the original corneal topography unchanged. To assess the performance, the accuracy of reconstructed conreal topographies before and after applying the filtering technique is studied. This is done by fitting the corneal height-data using Zernike polynomials of specific radial-orders and then observing the differences in the root mean square (RMS) errors. To account for limited data, resampling-based techniques are used for model-oder selection and the differences between the selected orders, before and after applying the filtering technique, are observed. These differences can be considered as an assessment tool for the accuracy and limitation of such a filtering technique.

The second aim is to compare a wide range of model-order selection criteria (both information theoretic- and resampling-based) to select the proper model-order for corneal surfaces from different clinical groups when they are fitted using orthogonal Zernike polynomials. The objective is to find a simple, but efficient, model-order selection criterion that can be relied on to select orders that agree with clinical expectations and at the same time to discriminate between different corneal groups with as-high-as-possible certainty. For this purpose, we propose to use the efficient detection criterion (EDC) due to the advantage it has over other criteria in terms of using *non-fixed* (flexible) penalty functions that usually control the performance of information theoretic-based model selection criteria. On the side and in order to compare different model-order selection criteria, simulations for theoretical regression polynomials are utilized. Biased estimation versus unbiased least squares estimation (LSE) are investigated in the context of model and model-order selection. Moreover, combining the searching methods (i.e., nested full-model search and partial-model search) for obtaining consistent model-order estimation is examined in order to use the advantage of the sparsity property that the model selection search method has.

1.2 Summary of the Contributions

The major contributions of this work are summarized as follows:

- Enhancement of videokeratoscopic images for more accurate corneal topography estimation [16]. Selecting of the proper parameters for the used filters (i.e., the adaptive pixel-wise Wiener filter and the morphological closing filter) leads to recovering the missing corneal data that resulted from non-standard image acquisition scenarios and/or irregularities in the rings' patterns due to mainly break-ups in the pre-corneal tear film. Moreover, and yet importantly, the proposed enhancement procedure does not change the original corneal topography.

- Using the efficient detection criterion (EDC) for model-order selection of corneal surfaces when fitted with Zernike polynomials [17]. The EDC is a simple and efficient criterion to be used for selecting the proper number of Zernike terms that need to be fitted to the corneal surface. The consistency of the EDC is the highest amongst all other tried criteria. Moreover, it shows the ability of discriminating between different corneal groups with a proper certainty.

- Improving the performance of model-order selection criteria using partial-model selection search [18].

- Applying biased estimation techniques leads to increasing the consistency of model and model order selection criteria [19].

- Using the hook-and-loop based model and model-order selection outperforms other criteria in terms of selecting the model-order for corneal surfaces when fitted with Zernike polynomials provided that the corneas are not highly irregular [20].

- Developing the hook-and-loop resampling plane and comparing its performance to other model and model-order selection criteria [21].

1.3 Overview of the Dissertation

Following this introduction, the dissertation is organized as follows:

Chapter 2: introduces the basics of corneal reconstruction and modeling. It highlights the important aspects related to the work achieved in this thesis. Details about the vision process, the cornea and its role in vision, basics of corneal topography and its applications, Placido disk videokeratoscopes and their applications, and corneal modeling using Zernike polynomials, are given.

Chapter 3: provides the details about the proposed procedure for enhancing videokeratoscopic images in order to obtain more accurate corneal topography estimation. Simulation and experimental results are provided in order to support the validity of using such an enhancement procedure [16].

Chapter 4: provides the details about using the EDC for model-order selection of corneal surfaces. The performance of the EDC was compared against the performance of a wide range of information theoretic- and resampling-based criteria. Simulation and experimental results are given to support the assumption that the EDC is capable of providing more consistent results in selecting the proper number of Zernike terms and, moreover, its ability to discriminate between different corneal groups [17].

Chapter 5: draws conclusions and provides outlook about the work of this thesis.

Appendix A: gives theoretical details about different model and model-order selection techniques. Moreover, simulation results were provided to show the superiority of using biased estimation for model and model-order selection [19] and how can the performance of model-order selection criteria (in particular bootstrap) improve significantly when partial-model selection search is used [18].

Chapter 2

Reconstruction and Modeling of Corneal Surfaces

2.1 Introduction and Motivation

In this Chapter, we focus on the corneal surface of the human eye and its important concepts related to the work achieved in this thesis. More comprehensive reviews can be found in [22–24].

Data from corneal surfaces are traditionally measured using a variety of techniques from which we concentrate on Placido disk based ones, particularly videokeratoscopes. Corneal data can be retrieved in many forms such as two dimensional (2D) videokeratoscopic (VKS) images from which the three dimensional (3D) points cloud (i.e., corneal height-data) can be reconstructed, and corneal slope-data from which the optical power of different zones of the corneal surface can be calculated. These forms of *corneal topography* measurements are important sources of information for screening and evaluating the normality (abnormality) of corneal surfaces.

Corneal height-data is usually modeled using orthogonal Zernike polynomials as each term of the modeling can be related to an optical property of the cornea. Therefore, we restrict ourselves to Zernike polynomials expansion of corneal surfaces.

2.2 The Vision Process

The human eye is a complex anatomical device where the ultimate goal is to allow humans to focus images on the retina. Fig. 2.1 illustrates a simple general structure of the human eye, where each optical element of the eye plays a distinct role in the vision process. The eye is an opaque ball filled with a water-like fluid. The transparent front surface of the eye is called the cornea. The cornea has the role of protecting the eye and refracting the light as it enters the eye. As parts of the outer surface of the eye, the cornea and the sclera maintain the intraocular pressure of the eye and resist infections.

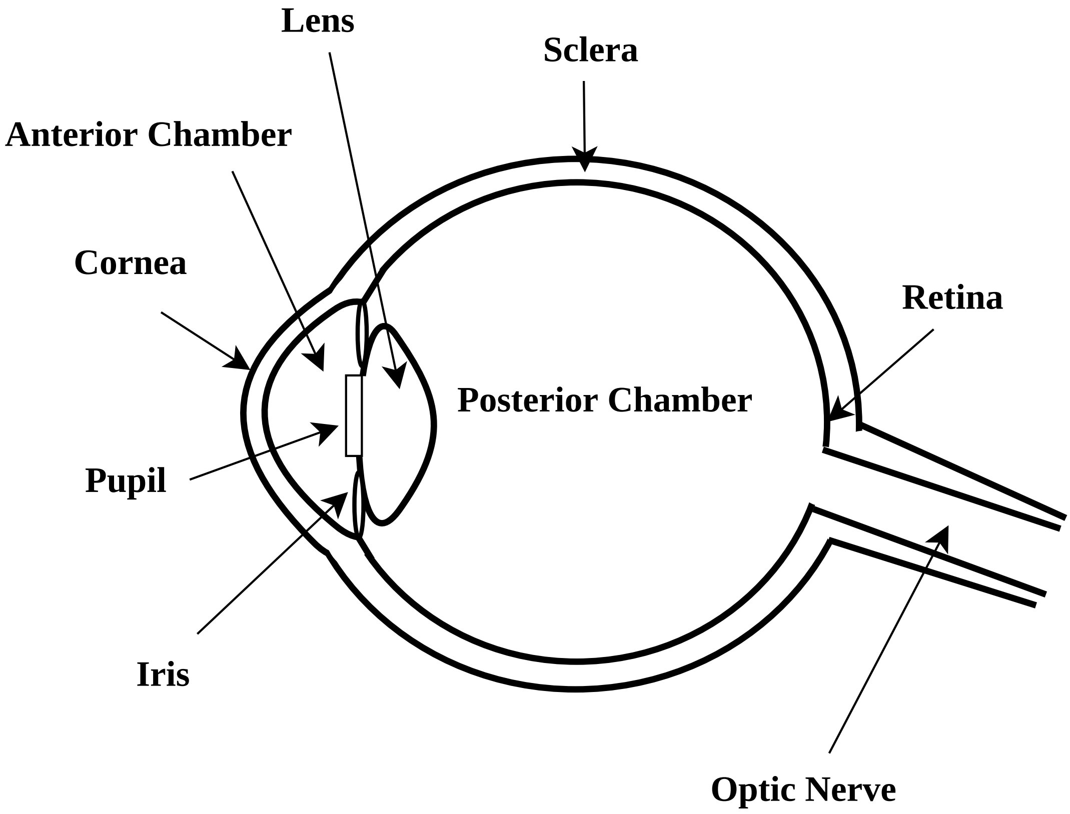


Figure 2.1. A simple general structure of the human eye.

After light passes through the cornea, a portion of it passes through an opening of the iris, the pupil. The iris, which is visible through the cornea, contains muscles which react to incoming illumination, and hence, the iris has an automatic ability to modify the pupil size playing an important optical function as an aperture. In bright-light conditions, the iris is minimized to reduce the size of the pupil and limit the amount of light which enters the eye and vice versa for dim-light conditions. Light refracted from the cornea enters the crystalline lens, which further focuses the light onto the back of the eye, the retina. The crystalline lens is connected to muscles that relax and contract in order to change the shape of the lens (and therefore its refractive power[1]) to allow the eye to focus on objects at different distances (i.e., accommodation). The retina, which is an extension of the central nervous system, contains light-sensitive cells that convert the light energy into neural signals sent to the brain for interpretation via the optic nerve.

There are some eye's models that approximate the typical eye to some degree [22].

---

[1] Refractive power (or optical power) of the eye is the ability of the eye to bend light as light passes through it. The eye that has much or little refractive power to focus light on the retina has a refractive error and, therefore, has a reduced visual acuity.

A simple approximation is called a reduced eye, in which the cornea and lens are replaced by a single optical surface having equivalent optical power of the cornea-lens combination.

2.3 The Cornea

The cornea is the major refracting component of the eye (the cornea is transparent and it is basically the only invisible outer part of the human body) and contributes to about 70% of the total optical (refractive) power of the eye, while the rest is provided by the lens [1, 2].

The mean refractive index[2] of the cornea is 1.376 [24]. There are no blood vessels in the cornea and, therefore, it is normally clear and has a shiny surface. However, the cornea is a highly sensitive surface because it has many nerve endings.

The normal cornea is a prolate surface (i.e., steeper in the center and flatter in the periphery). Its thickness varies from about 0.5 mm at the corneal apex (i.e., the geometric center of the cornea) to about 0.7 mm at the junction between the cornea and the limbus. Accordingly, there are differences in the radii of curvature between the posterior- and anterior corneal surfaces, as the posterior radius of curvature, R_p, is smaller than the anterior one, R_a. They are, however, approximately linearly dependent [25]

$$R_p = 0.81 R_a, \tag{2.1}$$

when the spherical assumption of the posterior- and anterior corneal surfaces is held.

The visually significant area of the corneal surface is approximately the area with the same diameter as the pupil size (which decreases with age). The central 4 mm of the cornea is approximately spherical [26]. Outside this optical zone, the cornea can be assumed in a first approximation to be aspheric[3] and radially asymmetric [26]. A smooth layer of tear-film covers the cornea, which is an essential component for the regular refraction of light by the eye, the formation of a clear retinal image, and optimal visual acuity.

The anterior surface of the cornea is the part that causes the largest refraction of the light entering the eye. This is mainly because it is the only air-to-tissue interface in

---

[2] The refractive index of a medium is a measure of the relative speed between waves in such a medium and a reference medium.

[3] An aspheric surface is slightly varying from a perfectly spherical shape.

the eye, and therefore causes a change in the refractive index that is much larger than anywhere else in the eye. The refractive power generated by the anterior corneal surface can be expressed in the paraxial region[4] as

$$P_a = \frac{n_a - 1}{R_a}, \tag{2.2}$$

where P_a is the refractive power in dioptres (D), n_a is the refractive index of the anterior corneal surface, and the value 1 is the refractive index of the air. For a normal cornea, the average value of the anterior radius of curvature is about 7.8 mm [26]. This contributes to +48.2 D, with $n_a = 1.376$.

Eyes are generally classified into three categories: emmetropic, myopic, and hyperopic. These categories are summarized as follows:[5]

1. Emmetropic eyes are those when relaxed (i.e., unaccommodated) focus distant parallel light rays quite well onto the retina. The emmetropic eye is considered as the "normal" eye.
2. Myopia, or nearsightedness, describes an eye that when unaccommodated and viewing a distant object, focuses the light in front of the retina. This causes distant object to be blurred out. Myopia dominantly occurs when the axial length of the eye is too long and/or the cornea is too steep.
3. Hyperopia, or farsightedness, is the opposite of myopia, in that the unaccommodated eye forms distant light rays behind the retina. Hyperopia dominantly occurs when the axial length of the eye is too short and/or the cornea is too flat.

On the other hand, there are optical defects (refractive errors) that are directly belonged to the corneal surface itself. We present here two types of optical defects that are related to the work achieved in Chapter 4:

1. Astigmatism is a common refractive error in which no single focus-point is formed due to the unequal refraction of light in different meridians. Therefore, no sharp image is focused on the retina and, consequently, a blurred vision results in. Corneal astigmatism occurs as a result of different curvatures between the two principal meridians of the anterior cornea [28] and means that the cornea has a higher curvature in one direction than the other. In most individuals, the

---

[4] Near the optical axis.

[5] More details are found in [27].

horizontal meridian is the flattest and the vertical meridian is the steepest, which can be linked to the difference in pressure caused by the upper and lower eyelids and known as with-the-rule corneal astigmatism. Irregular astigmatism is another type in which the corneal curvature is asymmetric along the same meridian and can be used to describe asymmetric aberrations (e.g., coma). However, exact causes of astigmatism are still unknown [28].

2. Keratoconus (or conical cornea) is an eye condition which results in the thinning of the cornea, as well as localized regions of high curvature. This almost-conical shape usually reduces the patient's visual acuity and is associated with high myopia and irregular astigmatism. The visual disturbance caused by keratoconus is mainly due to the irregular shape of the corneal surface.

There are topographic parameters (i.e., statistical indices) that summarize the optical quality of the corneal surface [29]. The surface regularity index (SRI) is a measure of the central corneal optical quality based on comparing the corneal powers of the adjacent points. Normal corneas have low SRI values while those with irregular astigmatism have higher values. Another parameter, the surface asymmetry index (SAI), is a weighted summation of differences in corneal power between the corresponding points 180° apart on equally spaced meridians. Normal corneas are highly symmetric and, therefore, their SAI are low. The SAI can be used for monitoring the progression of keratoconus where higher SAI values occur. Simulated keratometric values (simK)[6] and corneal power in the flattest corneal meridian (minK) are other measures that can be used for calculating the astigmatism [29].

The ability of the eye to resolve images and distinguish details is affected by refractive errors. These optical defects caused by the components of the eye (i.e., cornea and lens) degrade image quality, blur the image, and/or create a loss of visual acuity. Low-order refractive errors can be corrected with appropriate ophthalmic[7] lenses. This includes spectacles and contact lenses, which tend to refract the light rays before they hit the cornea [30]. However, high-order refractive errors can not be corrected with ophthalmic lenses. An alternative way to correct refractive errors of the eye is corneal refractive surgeries, which has become quite popular [31]. This surgical technique aims to change the shape of the cornea using laser beams in, for example, *laser-assisted in situ keratomileusis* (LASIK) surgeries to correct the vision. This means that the myopic cornea would be made flatter while the curvature of the hyperopic cornea should be increased by making it steeper.

---

[6] simK provides the corneal powers in the steepest meridian and the meridian perpendicular to it.

[7] Ophthalmology is the branch of medicine that deals with visual diseases and surgeries.

2.4 Basics of Corneal Topography

Corneal topography measurement is a non-invasive medical imaging procedure for representing the curvature of the corneal surface. The assessment of corneal topography is a valuable tool in the diagnosis and management of different corneal conditions. State-of-the-art corneal topography systems have begun in 1984 when computer analysis and image processing techniques were combined into the videokeratoscope [32]. Comprehensive details for historical background of the evolution of corneal topography can be found in [33–35].

The principle of measuring the corneal topography depends on calculating the anterior radius of corneal curvature, R_a, appeared in Eqs. (2.1) and (2.2). The basic formula falls withtin Gaussian approximation that assumes the cornea to be a sphere (i.e., a convex mirror) and it measures the size of an object reflected from the cornea. The concept is illustrated in Fig. 2.2 and the radius of curvature can be expressed mathematically as

$$R_a = 2D_o \frac{O_i}{O_o}, \tag{2.3}$$

where D_o is the distance between the object to the corneal vertex plan, O_i is the size of the object's image, and O_o is the size of the object. This method is adopted by the keratometer [36], which was the most popular clinical method for measuring the corneal topography until mid-1980s. However, the keratometer has a limited accuracy due to the spherical assumption of the cornea as well as the small coverage zone of the cornea it considers.

There are other methodologies for measuring the corneal topography, from which we summarize:

1. Raster photogrammetry [37, 38]. For such a method, a grid, called the raster stereographic grid, is projected onto a corneal surface covered with a fluorescein dye. The reflection of this grid is compared with a reference that is reflected from a flat surface and then the differences are used to compute the corneal topography. Even though this method yields a large coverage area of the corneal surface, using the fluorescein may change the properties of the cornea and, therefore, provides improper measurement of the corneal topography.
2. Slit scanning photography [39], which uses for example a slit-lamp[8] with a narrow beam. The beams are projected onto the corneal surface and a series of reflected

---

[8] A lamp that emits an intensive light's beam that enables an ophthalmologist to examine the retina and the optic nerve.

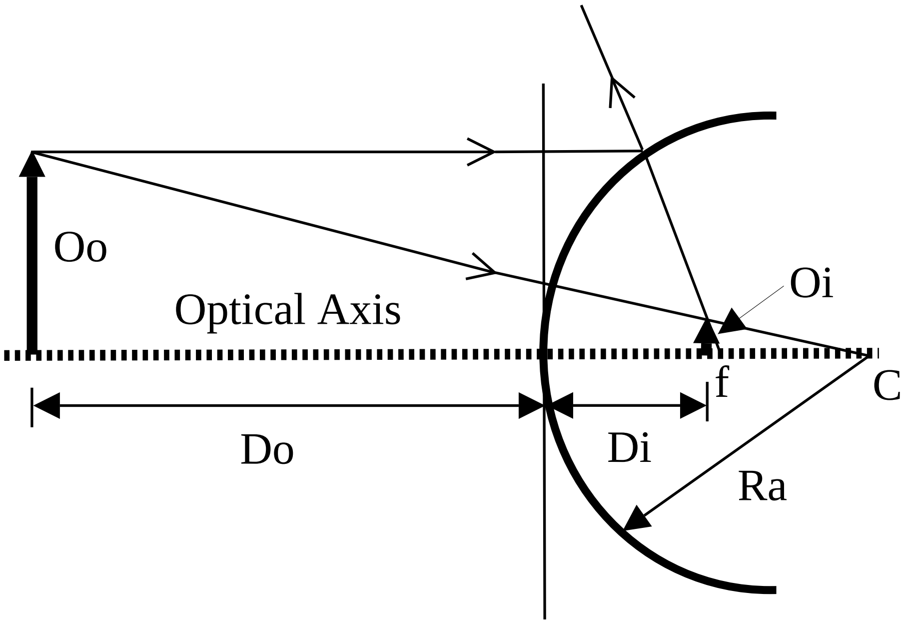


Figure 2.2. The basic concept for calculating the anterior radius of corneal curvature.

rays can be captured by a camera and then used to reconstruct the corneal shape employing the same principle of raster photogrammetry. The limitations of this technique are the time consumption and the artefacts introduced by eye movements.

3. Wavefront analysis with Hartmann-Shack wavefront sensor (HSS) [40] can be applied for measuring corneal topography. A laser beam is focused on the cornea by an objective (i.e., a lens) in such way that the center of the corneal curvature lies exactly in the focal point of the objective. If the corneal surface is perfectly spherical, the laser beam is reflected back onto itself at every point on the surface. If it is not the case, and the corneal shape varies from a perfect sphere, the laser beam is reflected in other directions on locations where there are differences. From these differences, the wavefront can be reconstructed, and then the corneal topography can be calculated.

The most common form of displaying corneal topography measurements is the color-coded power maps, in which each point on the map refers to a location on the corneal surface, and the corresponding color is a representation of the local dioptric power. Three different definitions of dioptric power are provided as follows [41, 42]:

1. Axial power, which describes the local corneal slope. The paraxial form of Eq. (2.2) is used to calculate the axial power by replacing R_a with the axial

distance, which is the perpendicular distance from the corneal surface to the optical axis.

2. Instantaneous power, which represents the local corneal curvature. Again, Eq. (2.2) is used to calculate the instantaneous power by replacing R_a with the instantaneous radius of curvature.
3. Refractive power, which uses Snell's law to describe the local refracting ability of the cornea [43].

It is worthy to mention that a direct relationship between axial and instantaneous powers was derived in [41].

Clinicians always need to know the shape and the refractive power of the cornea for a variety of applications. For example:

1. Fitting of contact lenses [44]. Knowing the corneal shape enables the clinician to have a better initial estimate to fit contact lenses more efficiently.
2. Planning of refractive surgeries (i.e., LASIK) [45]. Examining the corneal topography before and after refractive surgeries helps in evaluating the subject's treatment or, in some cases (e.g., keratoconic subjects), the subject's suitability for the treatment.
3. Monitoring the corneal changes in soft contact lens wearers (e.g., lens positioning from overnight wear) [30].
4. Corneal transplantation which can be a consequence from a failed refractive surgery [30].
5. Screening and diagnosing eye diseases that can affect the vision process (e.g., dry eye syndrome) [44].

2.5 Placido Disk Based Videokeratoscopy

Videokeratoscopy is considered as having a significant improvement in measuring corneal topography over keratometry and the other methodologies. It provides more details about corneal curvature variations over a large portion of the corneal surface.

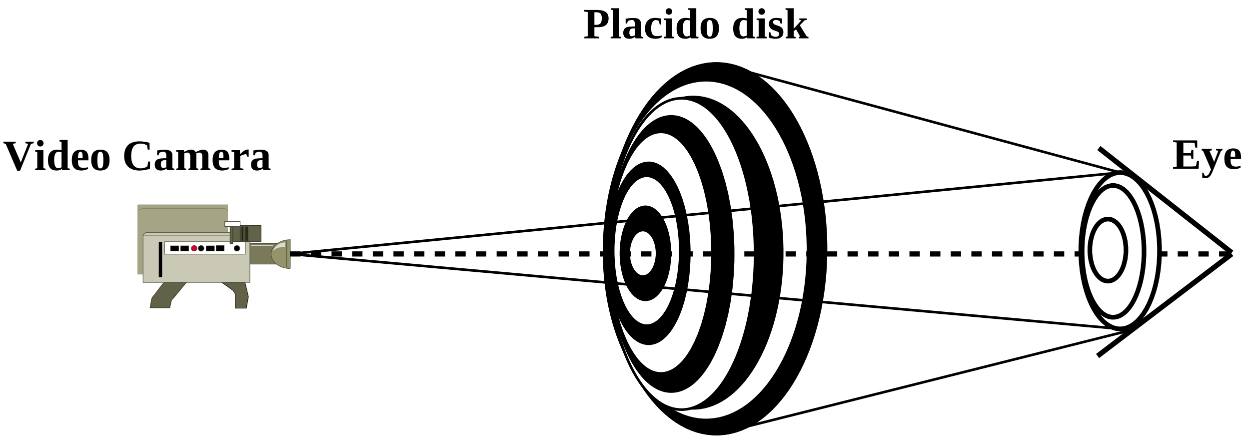


Figure 2.3. The concept of Placido disk based videokeratoscopes.

2.5.1 Basic Principle

Placido disk based videokeratoscopes contain a conical Placido disk target, an imaging system (an objective lens and a charge coupled device (CCD) video camera), and a computer system. The Placido disk target projects a number of illuminated and concentric rings on the corneal surface and a virtual image located behind the reflecting surface is detected by the video camera, that is positioned at the videokeratoscope's axis, and analyzed to determine the size and shape of the rings. This concept is illustrated in Fig. 2.3.

Each of the reflected rings is analyzed at a certain angular interval (e.g., 256 meridians) to allow computation of topography for a total of many thousand points. On the videokeratoscope's axis, there is a fixation point that is centered with respect to the concentric rings pattern. The position, size, and spacing of the reflected rings are determined by the corneal shape. If the cornea is spherical, the rings appear circular and regularly spaced. On the other hand, in astigmatic corneas, the reflected rings have an elliptical shape. Closer reflected rings indicate steeper corneas and, hence, larger refractive power than that of a sphere while further-apart rings yield flatter corneas and, therefore, smaller refractive power than that of a sphere. Hence, ring spacings in the resultant images are used to determine the local slope of the corneal surface. Moreover, local irregularities in the corneal surface are indicated by distortions in the rings pattern.

There are several algorithms for analyzing the rings pattern [46]. Local curvature algorithms are based on instantaneous radius of curvature along a particular meridian and essentially utilizes local changes in curvature to generate corneal topography.

Spherically biased algorithms for this analysis assume that the cornea is spherical, and small errors are introduced as a result. The arc-step algorithm [47–49] is a surface reconstruction algorithm, which is not spherically biased. It reconstructs corneal surfaces composed of several multiple arcs. The corneal shape is reconstructed by adding adjacent arcs sharing a common tangent where they meet [50, 51].

Besides the power maps used to describe corneal surfaces, corneal height-data[9] is another form of representing corneal topography in videokeratoscopy, where the separation between corneal points and a reference plane can be calculated. Corneal height-data are either calculated by direct measurements or by integrating the slope information of the Placido disk rings [52]. Moreover, power maps can be obtained by differentiating corneal height-data in the meridional (i.e., radial) direction. The first derivative gives the corneal slope (i.e., used to calculate the axial power) and the second derivative provides the corneal curvature (i.e., used with the corneal slope data to calculate the instantaneous power). Simply, conversion from one form of corneal topography data to another is possible through integration or differentiation.

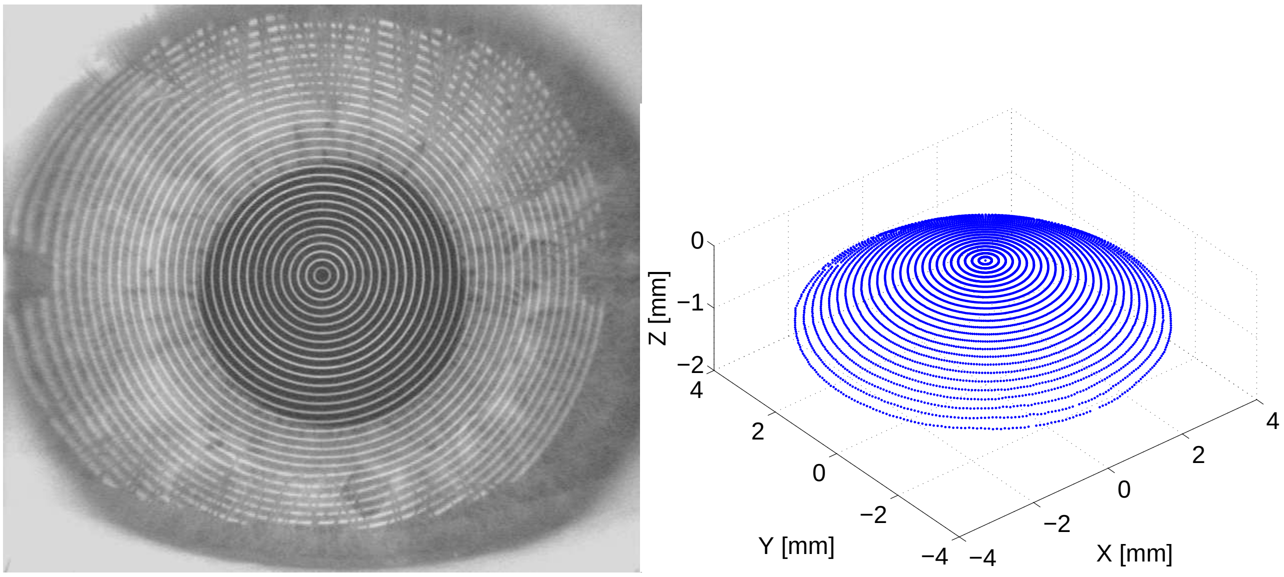
A number of commercial videokeratoscopes are found in the market. For most of them, the actual used algorithms are proprietary. For the work presented in this thesis, Medmont E300 (Medmont Pty Ltd, Melbourne, Australia) was used. Fig. 2.4 shows an example of a videokeratoscopic image for a normal cornea (captured by Medmont E300) along with its reconstructed corneal height-data. In [53], the performance of the Medmont E300 was compared to three other commercial videokeratoscopes. In terms of accuracy and precision, the Medmont E300 has shown good performance for aspheric test surfaces which resemble the real corneas.

2.5.2 Applications

Videokeratoscopes have been used for a wide range of clinical applications. A high-speed version of the Medmont E300 was developed [54] to enable data acquisition at a rate of 50 Hz. With such a high-speed videokeratoscope, more accurate analysis of the pre-corneal tear-film was possible as well as studying the dynamic response of the corneal surface to mechanical forces caused by the eyelids. In [55], high-speed videokeratoscope was also used for measuring the pre-corneal tear-film build-up time. In [56], the relation between the dynamic changes in the eye movements and the corneal shape was studied. There was a relationship between such movements and the pulse

---

[9] Corneal height-data is the source of corneal information used to achieve the work in Chapters 3 and 4.



(a) (b)

Figure 2.4. A videokeratoscopic image in (a) and its corneal height-data in (b) for an 8 mm diameter of a normal cornea.

and respiration signals, while there was no clear correlation between eye movements and corneal curvature.

In [57], an automatic algorithm for estimating the corneal limbus from videokeratoscopic images was proposed and its accuracy versus traditional manual techniques was demonstrated.

Estimating the corneal topography from videokeratoscopic image that have strong interference was done in [58], while recovering missing corneal areas for more accurate estimate of corneal topography was achieved in [16][10].

2.6 Corneal Modeling

Modeling of corneal height-data obtained from videokeratoscopy has gained significant interest amongst vision researchers since the introduction of videokeratoscopes. Fitting such data with parametric models yields an interpretable representation, which is in some cases better than the data itself. Such a modeling is appropriate for cross-sectional analysis of corneal topography [59], analysis of corneal response to accommodation [28],

---

[10] Full details are found in Chapter 3.

corneal classification [60], keratoconus detection [61], and high speed videokeratoscopy data compression [62].

Corneal modeling has ranged from simple shapes such as conic sections [63] to more complex models such as cosine hyperbolic functions [64] or the generalized conic functions [65].

Alternatively, corneal height-data has been modeled using Taylor series [66], radial polynomials [67], and spherical harmonics [68].

2.6.1 Zernike Polynomials

A dramatic drawback of viewing corneal height-data is that lower-order variations tend to mask higher-order variations, mainly due to the differences in coefficients' dynamic range. This problem can be overcome using Zernike polynomials expansion [69, 70].

The most popular modeling routine for videokeratoscopic height-data is a finite set of Zernike polynomials [4]. Such popularity of Zernike polynomial expansion lies in its relative simplicity, properties such as completeness and orthogonality over the unit circle, linear least squares estimation (LSE) of the parameters [71], and the optical properties of the expansion terms that can be linked to corneal aberrations [52].

Expanding the corneal height-data in cylindrical coordinates (ρ_n, θ_n, S_n), $n = 1, 2, \ldots, N$, using orthogonal Zernike polynomials can be expressed as

$$S(\rho,\theta) = \sum_{p=0}^{P} \sum_{\substack{q=-p \\ p-|q| \text{ even}}}^{p} a_p^q Z_p^q(\rho,\theta) + \varepsilon(\rho,\theta), \tag{2.4}$$

with $Z_p^q(\rho,\theta)$ being the Zernike polynomial with radial degree p and azimuthal frequency q, defined as [72]

$$Z_p^q(\rho,\theta) = \begin{cases} \sqrt{2(p+1)} R_p^q(\rho) \cos(q\theta), & q > 0 \\ \sqrt{2(p+1)} R_p^q(\rho) \sin(|q|\theta), & q < 0 \\ \sqrt{p+1} R_p^0(\rho), & q = 0 \end{cases} \tag{2.5}$$

where

$$R_p^q(\rho) = \sum_{l=0}^{(p-|q|)/2} \frac{(-1)^l (p-l)!}{l! \left(\frac{p+|q|}{2} - l\right)! \left(\frac{p-|q|}{2} - l\right)!} \rho^{p-2l}, \tag{2.6}$$

is the radial polynomial, a_p^q is the corresponding coefficient, and $\varepsilon(\rho,\theta)$ is the measurement and modeling error.

Alternatively, single-indexed Zernike polynomials [73] can be used for modeling corneal surfaces. In such a case, Eqs. (2.4) to (2.6) can be re-written as follows

$$S\left(\rho,\theta\right)=\sum_{k=1}^{K}a_k Z_k(\rho,\theta)+\varepsilon(\rho,\theta), \tag{2.7}$$

$$Z_k(\rho,\theta)=\begin{cases}\sqrt{2(p+1)}R_p^q(\rho)\cos(q\theta), & \text{even } k, q\neq 0\\ \sqrt{2(p+1)}R_p^q(\rho)\sin(q\theta), & \text{odd } k, q\neq 0\\ \sqrt{p+1}R_p^0(\rho), & q=0\end{cases} \tag{2.8}$$

where

$$R_p^q(\rho)=\sum_{l=0}^{(p-q)/2}\frac{(-1)^l(p-l)!}{l!\left(\frac{p+q}{2}-l\right)!\left(\frac{p-q}{2}-l\right)!}\rho^{p-2l}. \tag{2.9}$$

For simplicity, we adopt the notation of single-indexed Zernike polynomials in Chapters 3 and 4. Fig. 2.5 illustrates the first 10 Zernike polynomials according to [72] while Fig. 2.6 illustrates the first 16 Zernike polynomials according to the standard of single index [73].

2.6.2 Applications in Model-Order Selection

Model-order selection of corneal surfaces when expanded using Zernike polynomials has been used in many occasions. However, there is a continuous debate on the suitability of using such an expansion for efficient corneal modeling. For example, it was observed in [5] that Zernike polynomials do not represent the full details of highly irregular corneal surfaces, while they have shown, in [9], a better representation than other functional expansion methodologies in terms of reconstructing corneal surfaces.

Zernike polynomials with resampling-based model-order selection have shown superiority in model-order selection of normal corneal surfaces [6, 7], while it was found that they underestimate irregular corneal surfaces. This has been, however, overcome by introducing a refined bootstrap based model-order selection technique [8]. Nevertheless, the procedure is computationally expensive and needs many parameter settings. In [17], a simple and computationally efficient model-order selection criterion, the efficient detection criterion (EDC), has manifest itself as a good candidate to replace the refined bootstrap based approach[11].

---

[11] Full details are found in Chapter 4.

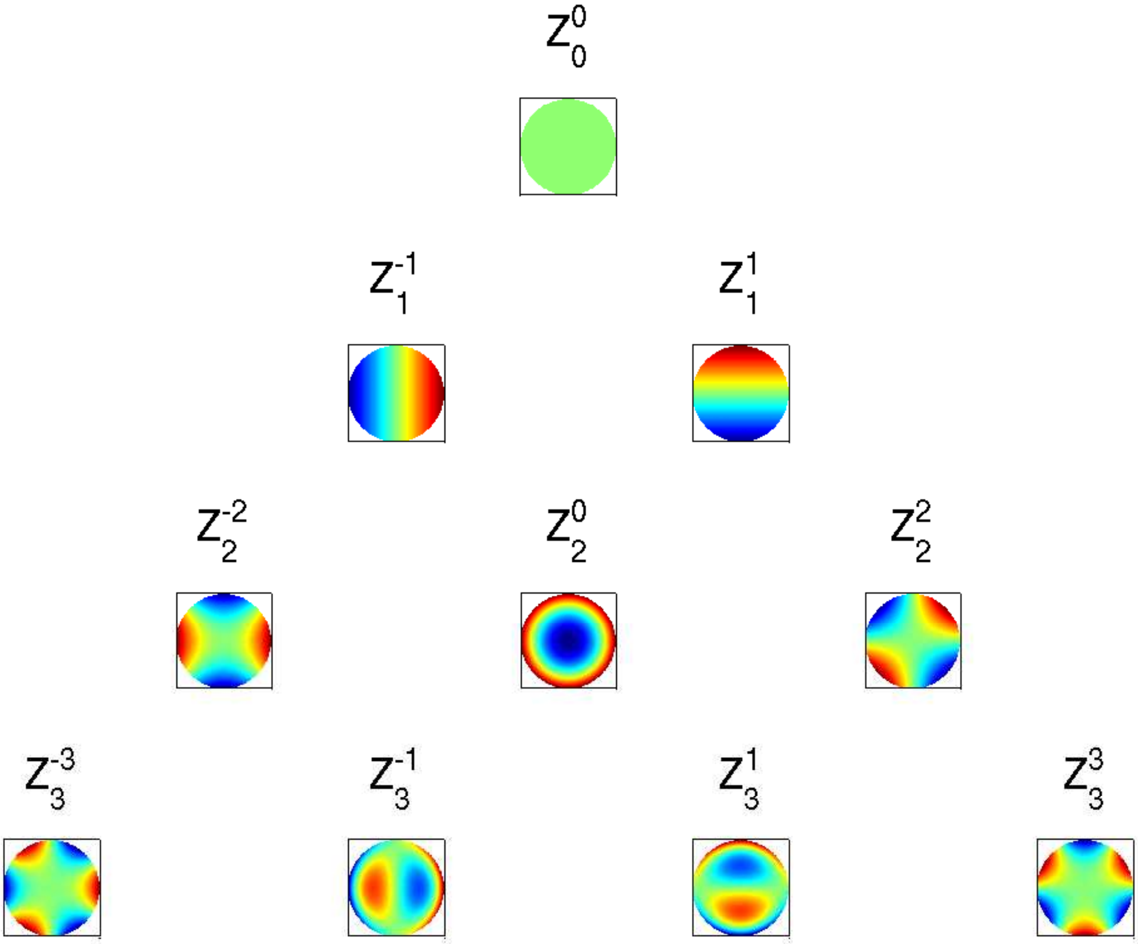


Figure 2.5. The first 10 Zernike polynomials according to [72].

2.7 Summary

In this Chapter, the main concepts related to corneal data representation was introduced. Steps of the image formation in the human eye were summarized, followed by a focus on the role of the cornea on the vision process. Basics of corneal topography, its displays, and applications were highlighted. As being the state-of-the-art methodology, the principle of Placido disk videokeratoscopes was given along with the corresponding applications. Zernike polynomials expansion was adopted as a fitting routine to corneal height-data in the work of this thesis. Therefore, using such a modeling procedure for corneal surfaces was detailed.

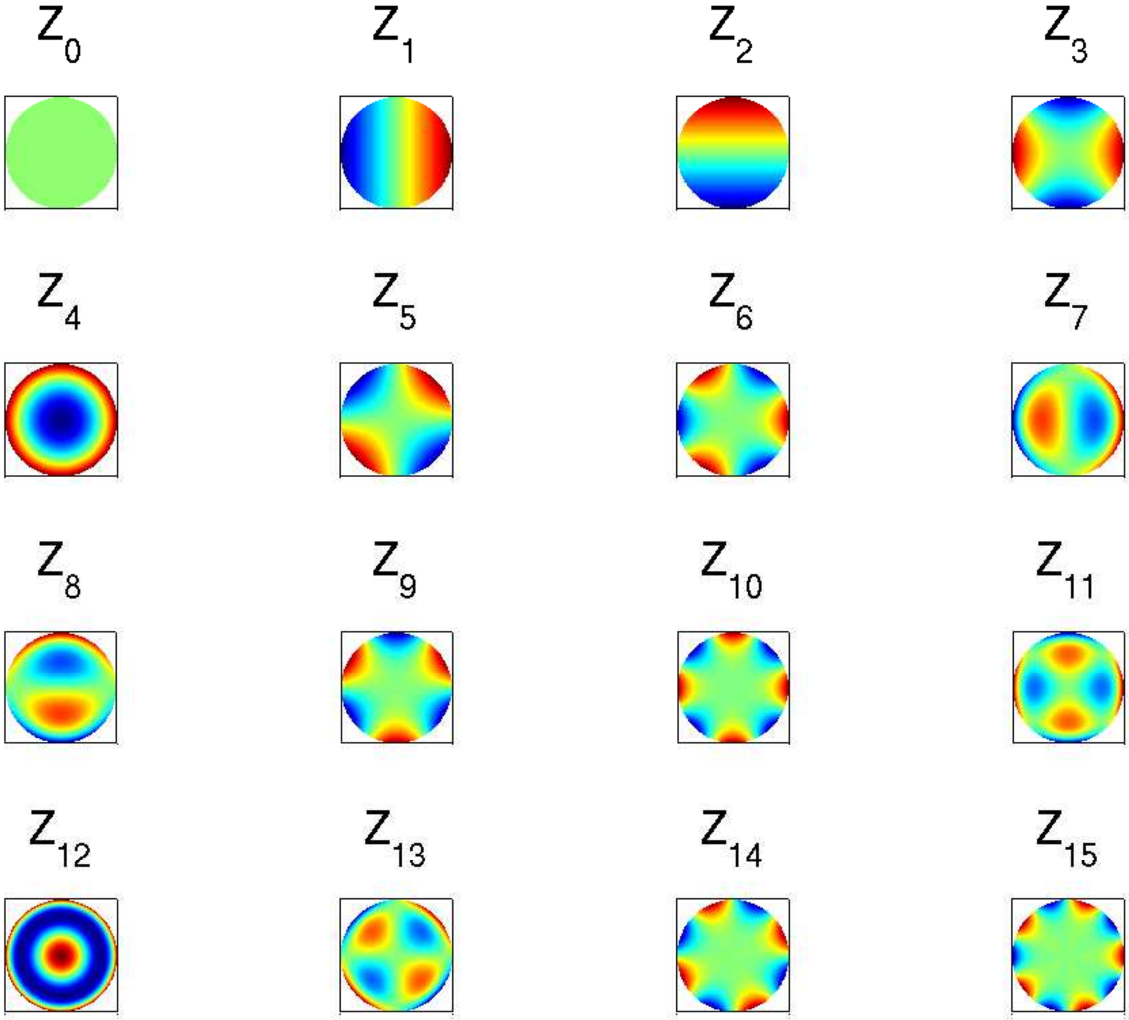


Figure 2.6. The first 16 Zernike polynomials according to the standard of single index [73].

Chapter 3

Enhancement of Videokeratoscopic Images

3.1 Introduction and Motivation

The current state-of-the-art of measuring the anterior surface of a cornea comprises instrumentation that is based on the projection of the Placido disk rings onto the cornea and registration (imaging) their reflection from the pre-corneal tear film surface [35]. The technique is referred to as videokeratoscopy or videokeratography as the imaging system is usually a charge coupled device (CCD) camera linked to a computer system that runs an algorithm for corneal topography estimation [50].

Many studies have been undertaken to evaluate the accuracy of the Placido disk based corneal topography reconstruction. They included theoretical error analysis [51, 74], issues such as the skew ray ambiguity [75, 76] as well as the performance evaluation of commercially available systems using test surfaces [53, 77] and a range of real corneas [78, 79]. However, much less attention has been given to improving the image processing routines [14, 58, 80].

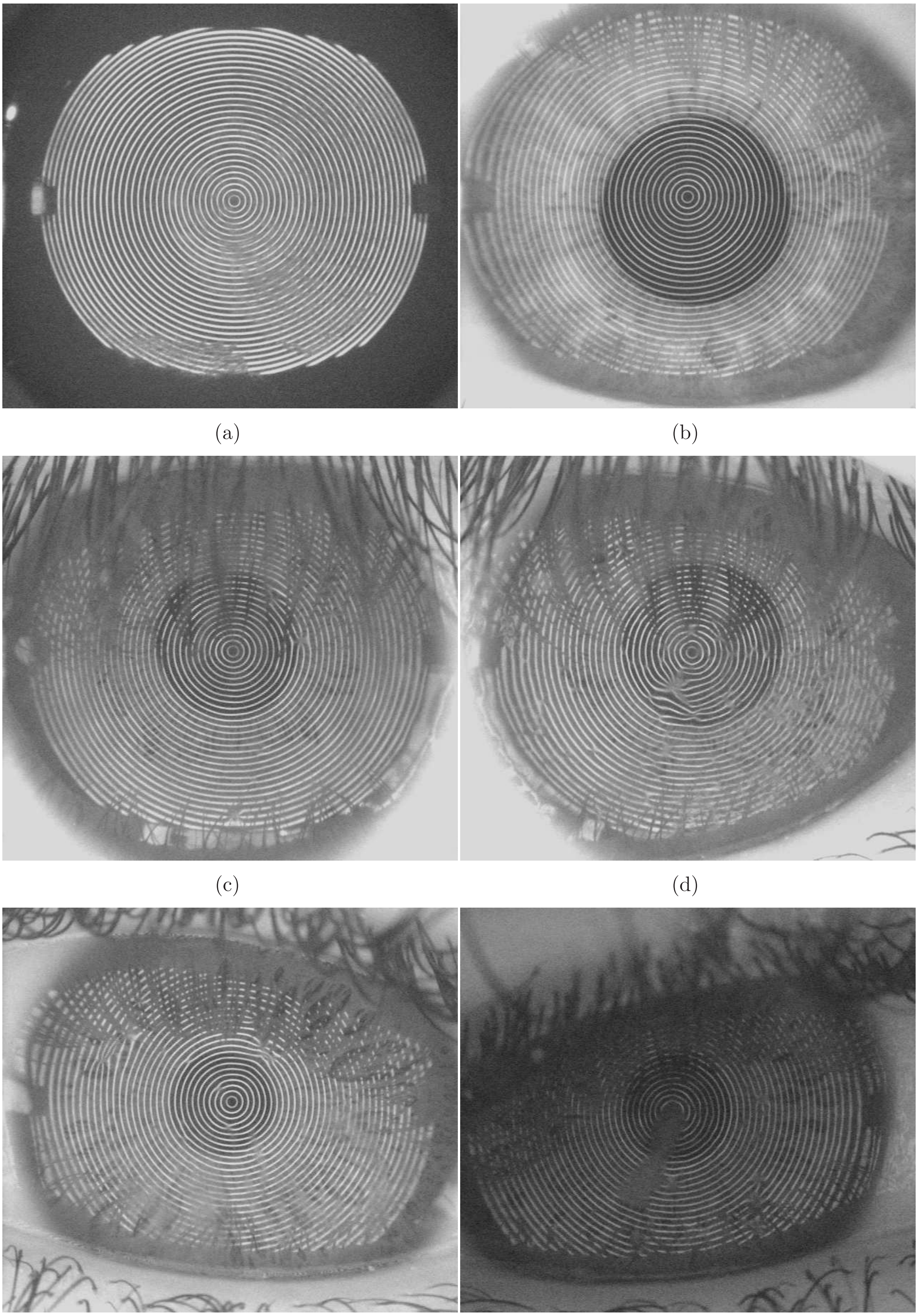
There are three parameters of videokeratoscopes that are of particular importance to clinicians. First, the accuracy of corneal topography reconstruction (normally evaluated with a help of calibrating sphere), which should be of order of a micron or less. Second, corneal coverage (expressed as a maximum corneal diameter), which is ideally covering the white-to-white corneal diameter [81]. Third, the ability to provide accurate estimates of real corneal topographies for a range of *non-ideal* conditions in which the subject fixation, eye movements, poor tear film quality, presence of mucus, and reflection from eyelashes (particularly in the superior cornea) contribute to a much poorer quality of the acquired videokeratoscopic image.

In the clinical setting the subject is typically requested to blink prior to a topography measurement in order to refresh the pre-corneal tear film and attempt to achieve a widely open ocular aperture. An operator usually manually aligns a videokeratoscope, waits about two seconds after the blink for the tear film to build-up [55], and acquires the videokeratoscopic image. Some commercially available videokeratoscopes have the ability to acquire an image automatically when in focus. To do this, they use a range finder that is embedded in the videokeratoscopic Placido disk bowl or cone. The instrument employed in the current study (Medmont E300, Medmont Pty Ltd, Melbourne, Australia) has this feature. This instrument automatically provides an "acceptance rating" (0-100%) of how acceptable an acquisition is. This rating is based on a range of criteria that includes the estimated distance between the CCD and corneal apex, as well as eye movements, and instrument centration. Only acquisitions in which a rating of $\geq 95\%$ were analyzed in this study. Fig. 3.1 shows some examples of standard videokeratoscopic images. The mentioned clinical setting ensures nearly ideal conditions for image acquisition and/or processing. Representative examples for such ideal conditions are shown in Figs. 3.1(a) and 3.1(b) for an 8 mm radius polymethyl methacrylate (PMMA) model sphere and a real eye, respectively.

However, there are numerous cases in which the acquisition is not ideal. Many subjects have difficulties opening their eyes wide. Some, like Asian subjects have narrow palpebral aperture (the distance between the eyelids), which normally leads to interferences from eyelashes' reflections as shown in Fig. 3.1(c), particularly in the superior part of the corneal surface. Tear film instability is yet another form of non ideal acquisition. The tear film is constantly changing. After a blink, it undergoes a formation (build-up) followed by a stable phase and a subsequent deformation (break-up) [82]. In subjects diagnosed with dry eye [83], the stable phase that is required for acquisition of the corneal topography is either very short or may not exist at all (i.e., the build-up phase is immediately followed by the break-up phase). A representative example where tear film break-ups are encountered is shown in Fig. 3.1(d). Moreover, in some cases, a significant part of the corneal surface may be covered by mucus as illustrated in Figs. 3.1(e) and 3.1(f). In all such non ideal cases, there are missing observations in the sampled corneal topography that need to be interpolated. When a large number of missing points are encountered, the area is usually omitted leading to a smaller corneal coverage. Areas of interference can be separated in a videokeratoscopic image from the useful Placido disk information as it was recently shown in [58]. That approach results in better corneal topography estimates but also in a much smaller coverage area.

The main goal of the work in this Chapter is to develop an improved image processing technique that overcomes these errors of missing or incomplete data in order to improve the estimation of corneal topography.

This Chapter is organized as follows: Section 3.2 presents the methodology of enhancing the raw videokeratoscopic image while Section 3.3 gives a description of the performance evaluation based on resampling techniques. Experimental results with discussion are provided in Section 3.4. A summary of this work is stated in Section 3.5.



(a) (b) (c) (d) (e) (f)

Figure 3.1. Examples of standard videokeratoscopic images: (a) Spherical model, (b) Standard acquisition, (c) Long eyelashes, (d) Tear-film break-ups, (e) Mucus, and (f) Severe mucus.

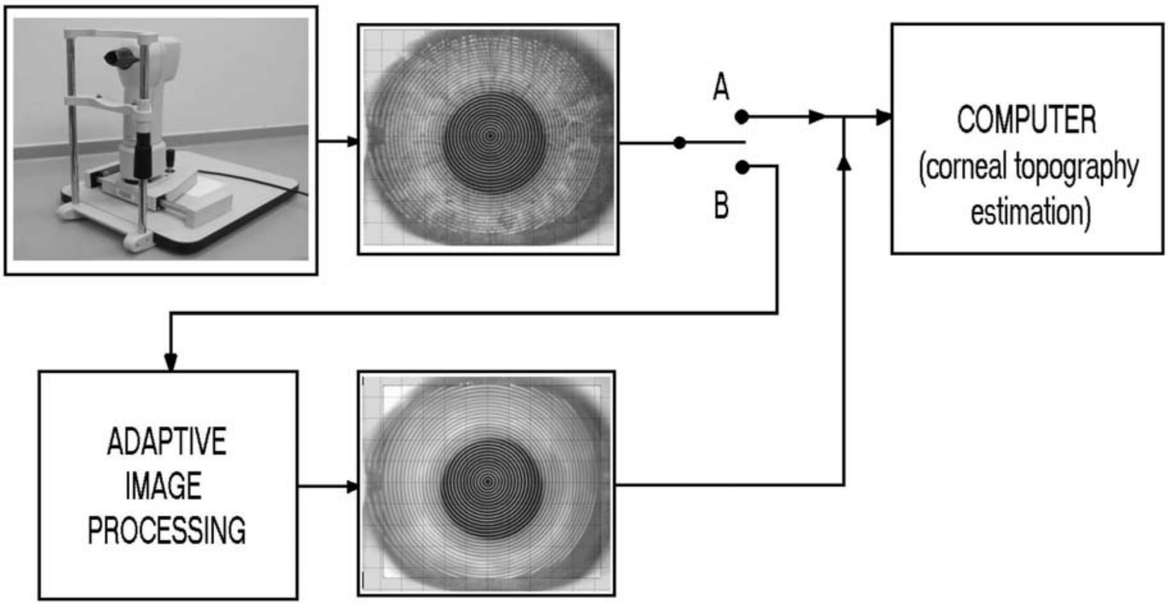


Figure 3.2. The concept of evaluating the performance of the proposed technique.

3.2 Methodologies

A limited number of studies has been published on the image processing techniques used for extracting the ring patterns in Placido disk ring images and there are no well-established algorithms, whose performance can be compared with any newly developed method. For example in [14], a graph-search based approach was used for labelling points of each ring. However, there was no consideration for the missing observations in the images. In [15], an approach based on the global coherence of the images was proposed but no real results were reported. Thus, a feasible way to evaluate the performance of any new technique for processing videokeratoscopic images is to use the instrument's own procedures (often being proprietary knowledge) for estimating corneal topography. This concept is illustrated in Fig. 3.2. Here, we employ a pre-processing technique comprised of adaptive image processing techniques to improve the Placido disk image (e.g. reduce the number of data outliers), and then this pre-processed image is fed back through the instrument's inherent processing algorithm, and the results of these two methods are compared. The reduction of outliers aims to improve the data quality and potentially increase the area of measurement coverage.

Fig. 3.3 summarizes the steps that form the block of adaptive image processing. Basically, the procedure starts by detecting the centroid of the Placido disk image and uses it for transforming the image from Cartesian to the polar coordinates. Processing the image in polar coordinates is easier than in the Cartesian domain because the original

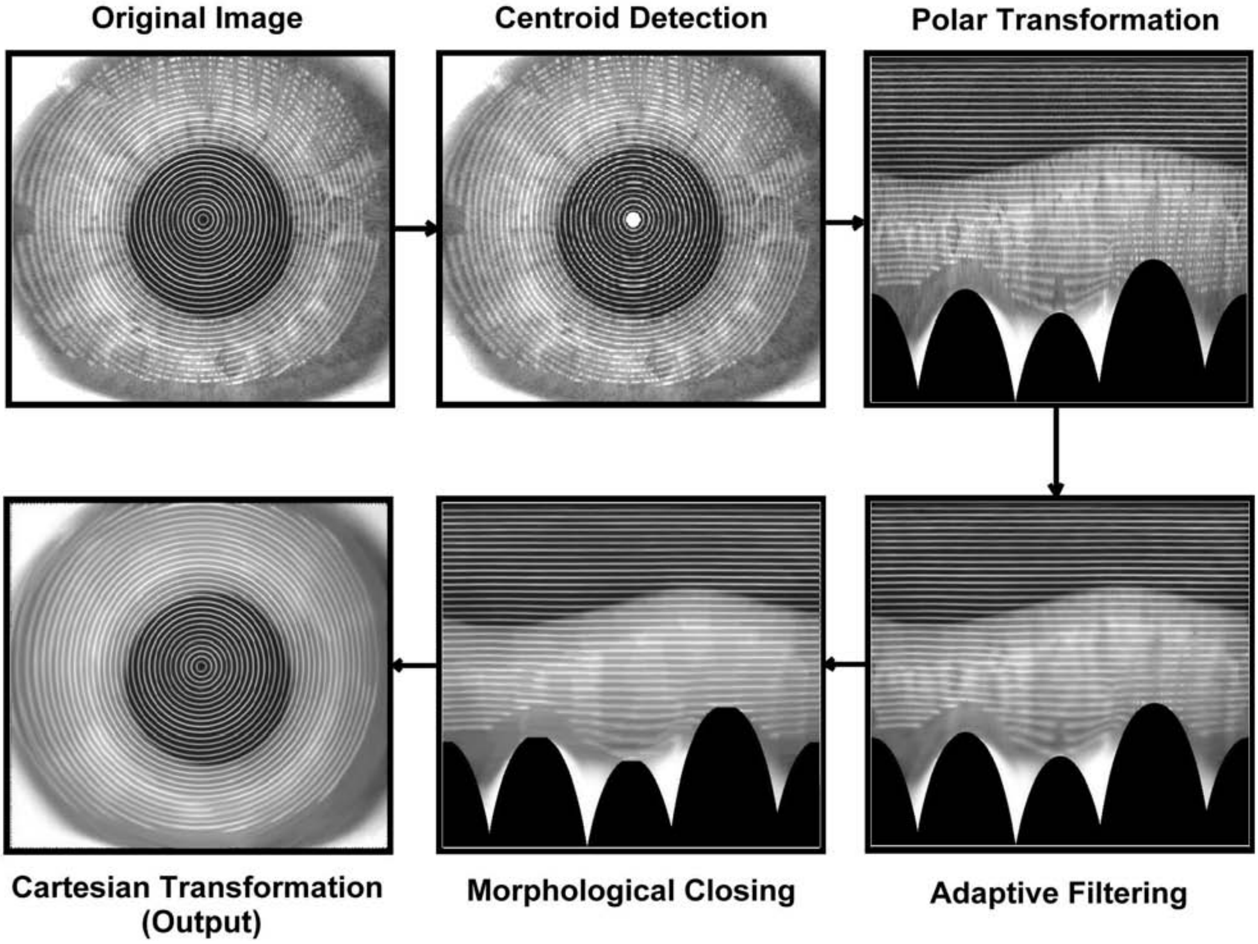


Figure 3.3. The steps of the adaptive image processing block.

quasi-circular and concentric Placido disk rings tend to form quasi-straight lines in the polar domain. The core of the procedure is the combination of the adaptive filtering (for smoothing Gaussian-like noise found after the image acquisition stage and at the same time preserving the edges of the rings) and the morphological closing operation (for removing the outliers). An inverse transformation, from the polar to Cartesian coordinates, is then applied to revert the processed image to its original domain.

To statistically assess the performance of the proposed methodology, the root mean square (RMS) error between the corneal height-data and its parametric modeling (i.e., using Zernike polynomial expansion of specific orders) is measured and compared between the original and the pre-processed videokeratoscopic images. Moreover, bootstrap-based techniques [10, 12, 13] are used to select the optimal model-order of the Zernike polynomial expansion to corneal height-data. The bootstrap mean-square error of the selected optimal-order (for the original and the pre-processed images) is measured and then used as an performance-evaluation indicator for the validation of

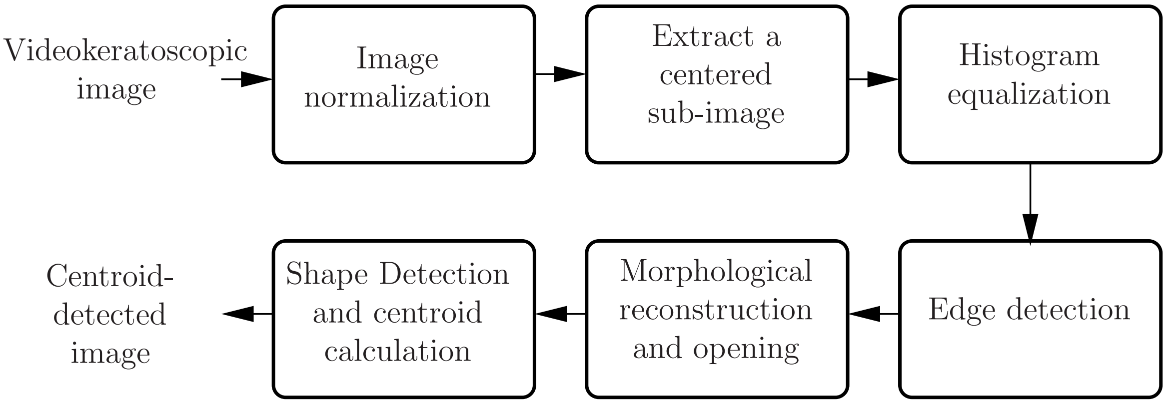


Figure 3.4. The steps of the centroid-detection algorithm.

the proposed adaptive image processing. However, more details about model-order selection in the context of corneal surfaces modeling will be thoroughly discussed in Chapter 4 utilizing both information theory- and resampling-based techniques. Next, a detailed description of each step of the proposed image processing algorithm is provided.

3.2.1 Centroid Detection

The videokeratoscopic image, denoted here by $I(x, y)$, is usually deteriorated by outliers, which are in the form of Gaussian-like noise, ring breaks (due to eyelashes reflections), and possible breaks caused by tear-film break-up phase. First, $I(x, y)$ is normalized

$$I_N\,(x, y) = \frac{I(x, y) - \min\{(I(x, y)\}}{\max\{(I(x, y)\} - \min\{(I(x, y)\}}, \tag{3.1}$$

so that the gray-levels of $I_N\,(x, y)$ are in the interval $[0, 1]$.

$I_N\,(x, y)$ can be further processed in the polar domain easier than in the Cartesian domain. Therefore, the centroid of the innermost ring must be detected in order to be used for precise polar transformation. However, finding this centroid is not straightforward, particularly when large areas of the rings are broken (e.g., as seen in Fig. 3.1(f) where mucus covers a large area of the central part of the cornea). The region of interest (centroid area) can be approximately found in the central region of the videokeratoscopic image. Accordingly, The main idea is to extract a squared central subimage of $I_N(x, y)$, find either a complete circular region or an arc of considerable length, and then compute its geometrical centroid. Fig. 3.4 summarizes the steps of the centroid-detection algorithm, while the detailed steps are as follows:

1. Define a squared sub-image, $J_S(i,j)$, whose center is $(n_x/2, n_y/2)$, where n_x and n_y are the number of pixels of $I(x,y)$ in x- and y- directions, respectively. $i = n_x/2 - M, n_x/2 - M + 1, \ldots, n_x/2, \ldots, n_x/2 + M - 1, n_x/2 + M$ and $j = n_y/2 - M, n_y/2 - M + 1, \ldots, n_y/2, \ldots, n_y/2 + M - 1, n_y/2 + M$, where $M = 50, 75, 100, \ldots, \min\{(n_x/2, n_y/2)\}$ determines the size of $J_S(i,j)$. The algorithm starts with $M = 50$ to minimize the searching time.

2. Perform a sequence of elementary image processing operations on $J_S(i,j)$ in order to locate the region that contains the searched centroid. They are

   (a) Histogram equalization [84] is used to stretch the contrast information in $J_S(i,j)$. This is a transformation utilizes the cumulative distribution function (CDF) of the gray-levels of $J_S(i,j)$ to calculate the gray-levels of the output image $J_E(i,j)$

   $$g_u = \sum_{v=0}^{u} \frac{N_v}{N_T} \qquad u = 0, 1, 2, \ldots, L-1, \tag{3.2}$$

   where N_v and N_T are the numbers of pixels with gray-level v and the total number of pixels in $J_S(i,j)$, respectively, and L is the total number of possible gray-levels. Therefore, the histogram of $J_E(i,j)$ is uniform. Experimentally, it was found that this step helps in giving more edge contours, which are required for image reconstruction (i.e., region filling) as explained below.

   (b) Edge detection is used to map $J_E(i,j)$ to an image of edge contours, $J_{BW}(i,j)$. Canny edge detector [85] is used for this purpose. It uses the first derivative of a 2D Gaussian function, $\Omega(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$, to calculate the gradient of the smoothed image (by a 2D Gaussian filter). This is followed by a non-maximum suppression stage to find local maxima in the gradient direction and suppresses all others (minimizing false edges). Then, a hysteresis thresholding is used by defining two thresholds, t_H and t_L where $t_H > t_L$. Gradient pixel values above t_H are immediately classified as edges. By tracing edge contours, neighboring pixels with gradient less than t_H can be still marked as edges once they are above t_L. This facilitates identifying strong edges, preserving relevant weak edges, and maintaining some level of noise suppression. This preprocessing step simplifies the process of finding circular or arc objects.

   (c) Morphological reconstruction [86] introduces the concept of geodesic, or equivalently conditional morphological operators (e.g., geodesic dilation that can be used for region filling). In a binary image, this can be regarded as filling holes (set of zeros surrounded by a boundary of ones) into the image. Mathematically, region filling is expressed as

$$X_\varphi = (X_{\varphi-1} \oplus SE(i,j)) \cap J_{BW}(i,j)^c, \quad \varphi = 1, 2, \ldots, \tag{3.3}$$

where $SE(i,j)$ is a symmetric structuring element, $J_{BW}(i,j)^c$ is the complement of $J_{BW}(i,j)$, and X_0 is selected as a point inside the boundary. This procedure is iterative and it stops at iteration φ if $X_\varphi = X_{\varphi-1}$. The filled-region image, $J_F(i,j)$, is expressed as

$$J_F(i,j) = X_\varphi \cup J_{BW}(i,j), \tag{3.4}$$

which includes filled circular and/or arc objects that are found to be necessary for calculating the centroid.

(d) Morphological opening [87] is mathematically expressed as

$$J_O(i,j) = J_F(i,j) \circ SE(i,j) = (J_F(i,j) \ominus SE(i,j)) \oplus SE(i,j), \tag{3.5}$$

which is an erosion of $J_F(i,j)$ by the structuring element $SE(i,j)$, followed by a dilation of the result by $SE(i,j)$. Conceptually, morphological opening is used to filter out the image objects that can not be fitted in the structure element. The goal here is to eliminate the remaining trivial edge contours (which could not be filled out) of $J_F(i,j)$.

(e) Calculating the area of the largest object in $J_O(i,j)$, simply by counting its pixels. If this area is relatively large, then the centroid can be calculated and the algorithm proceeds with the next step (step 3). Otherwise, the algorithm restarts from step 1 by increasing the value of M.

3. $J_O(i,j)$ is supposed to have only one relatively large circular or arc object. To automatically decide about the geometrical shape of this object, the eccentricity; which is mathematically defined as

$$E = \sqrt{1 - \frac{A_b^2}{A_a^2}}, \quad 0 \leq E \leq 1, \tag{3.6}$$

is used. Eccentricity is a parameter associated with every conic section[1] and can be used as a measure of how much the conic section deviates from being circular, where A_a and A_b are the semimajor and semiminor axes ($A_b \leq A_a$) of the conic section, respectively. The procedure of calculating the eccentricity [88] is performed by finding an ellipse that has the same second central moments as the circular/arc object, computing A_a and A_b, and then E. The eccentricity of

---

[1] A conic section is the curve produced by the intersection between a cone and a plane.

a circle is zero (because its radius is constant, and hence $A_a = A_b$) while the eccentricity of an arc is approximately one (because $A_b \ll A_a$ in this case). The algorithm decides about the geometrical shape of the object as follow:

(a) If E is small, then the object is approximately circular. Calculating the centroid is simply performed by averaging the locations of the points in the object (center of gravitation). For example, if the object has h points with coordinates; $(x_1, y_1), \ldots, (x_h, y_h)$, then the centroid is defined as:

$$(C_x, C_y) = \frac{1}{h} \sum_{i=1}^{h} (x_i, y_i). \tag{3.7}$$

(b) If E is approximating one, then the object is considered as an arc, with a possibility of being a thick arc. In the later case, the arc should be reduced to one-pixel-thick arc whose points can be used efficiently to fit a circle and then solving for its centroid by utilizing the least-squares approach. Morphological thinning [87] is the proper operator, which can be interpreted in terms of hit-or-miss transformation (a basic tool for shape detection) [87]. Hit-or-miss transformation looks for occurrences of particular binary patterns (objects) in an image. For the application at hands, it is sufficient to mathematically express hit-or-miss transformation as a simple erosion operation of the image of the object, $J_O(i,j)$, with a structuring element, $SE(i,j)$, which consists of the binary pattern to be detected

$$J_O(i,j) \circledast SE(i,j) = J_O(i,j) \ominus SE(i,j). \tag{3.8}$$

Morphological thinning is an operator that is used to erode selected pixels from the object pattern until only one-pixel-thick object remains. Mathematically, it is expressed as

$$\begin{aligned} J_O(i,j) \otimes SE(i,j) &= J_O(i,j) - (J_O(i,j) \circledast SE(i,j)) \\ &= J_O(i,j) - (J_O(i,j) \ominus SE(i,j)) \end{aligned} \tag{3.9}$$

In words, thinning operation is calculated by translating the origin of $SE(i,j)$ to each possible pixel position in the object, and at each such position comparing $SE(i,j)$ to the underlying object pixels. If there is an exact match between the pixels in $SE(i,j)$ and the pixels in the object, then the object pixel underneath the origin of $SE(i,j)$ is set to background (zero). Otherwise, it is left unchanged. Usually, a set of structuring elements, $SE(i,j) = \{SE_1(i,j), SE_2(i,j), \ldots, SE_n(i,j)\}$, are defined. Then, the image is thinned sequentially by each element (i.e., the result of thinning by $SE_1(i,j)$ is thinned by $SE_2(i,j)$ and so on). The whole process is repeated until no more changes in the output occurs. We have used one of the algorithms described in [89] to thin the detected arc object. The least-squares approach [71] can be used to fit a circle to the points of the thinned arc, by minimizing the sum of squared algebraic distances. It is well known that the general bivariate quadratic equation can be expressed as

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \quad a, b, c \neq 0 \tag{3.10}$$

For the circle case, this can be reduced to

$$ax^2 + ay^2 + dx + ey + f = 0. \tag{3.11}$$

If the arc object has h points $((x_1, y_1), \ldots, (x_h, y_h))$, then this linear model can be expressed in matrix form

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_h & y_h & 1 \end{bmatrix} \begin{bmatrix} d/a \\ e/a \\ f/a \end{bmatrix} = - \begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ \vdots \\ x_h^2 + y_h^2 \end{bmatrix}. \tag{3.12}$$

and then solving for the coefficients d/a, e/a, and f/a in the least-squares sense. Manipulating Eq. (3.11) by completing the square, yields

$$\left(x + \frac{d}{2a}\right)^2 + \left(y + \frac{e}{2a}\right)^2 = \frac{d^2 + e^2}{4a^2} - \frac{f}{a}. \tag{3.13}$$

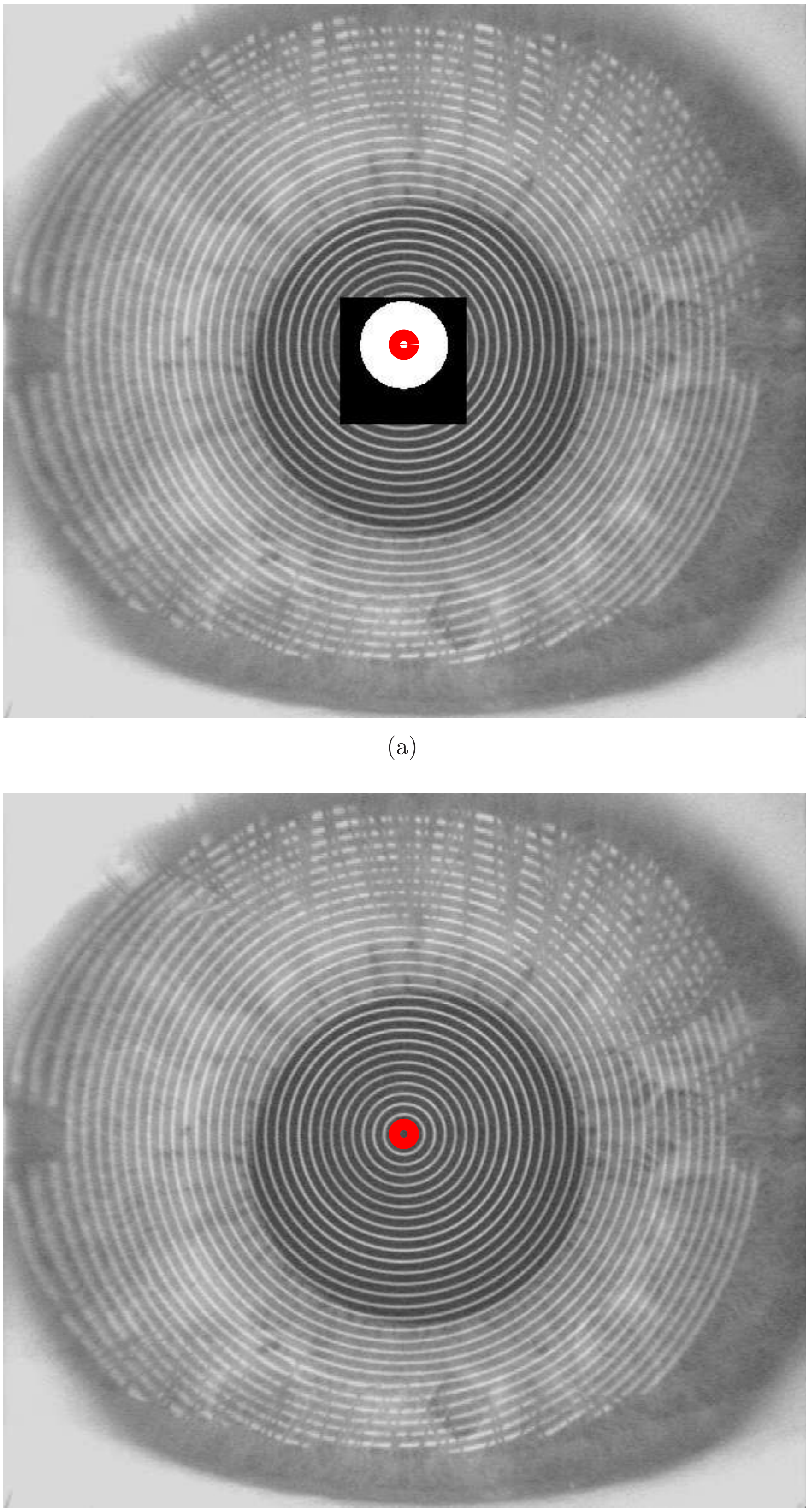
Eq. (3.13) represents the standard circle equation, from which the center can be expressed as

$$(C_x, C_y) = \left(\frac{-d}{2a}, \frac{-e}{2a}\right). \tag{3.14}$$

Figs. 3.5 and 3.6 show two examples for detecting the centroid of the Placido disk pattern when a complete circular object is found (in a standard-acquisition case) and when an arc object has been detected (in an VKS image with a large central area of mucus), respectively.

3.2.2 Polar Transformation

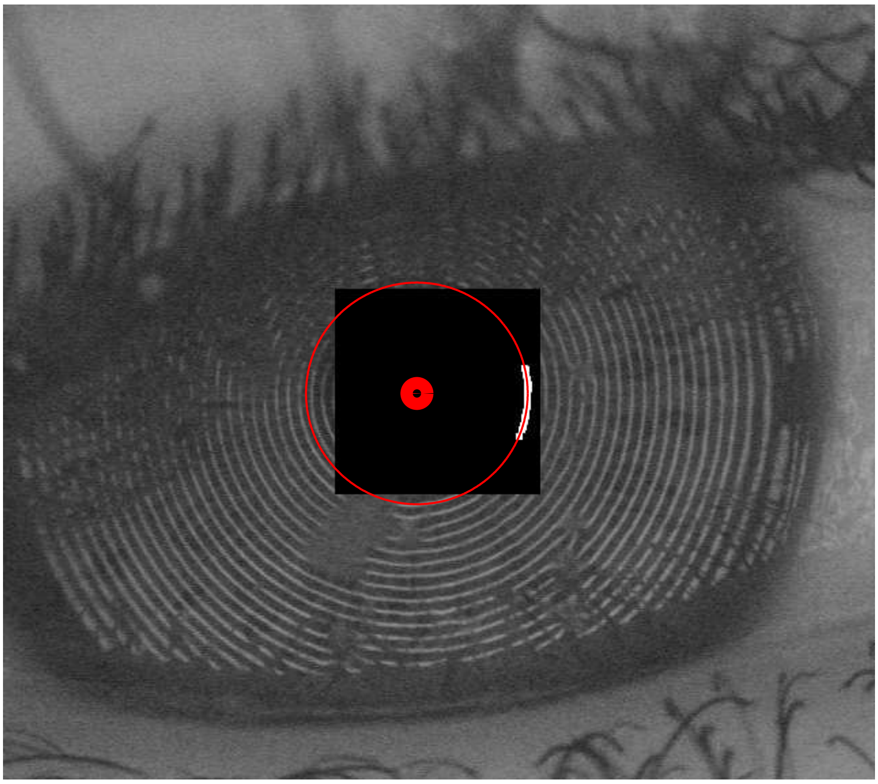
It is quite useful to transform the normalized VKS image, $I_N(x, y)$, to the polar domain using the detected centroid (C_x, C_y). The aim behind polar representation is to get an image of quasi-lines, in a specified direction (i.e., the horizontal direction here), which



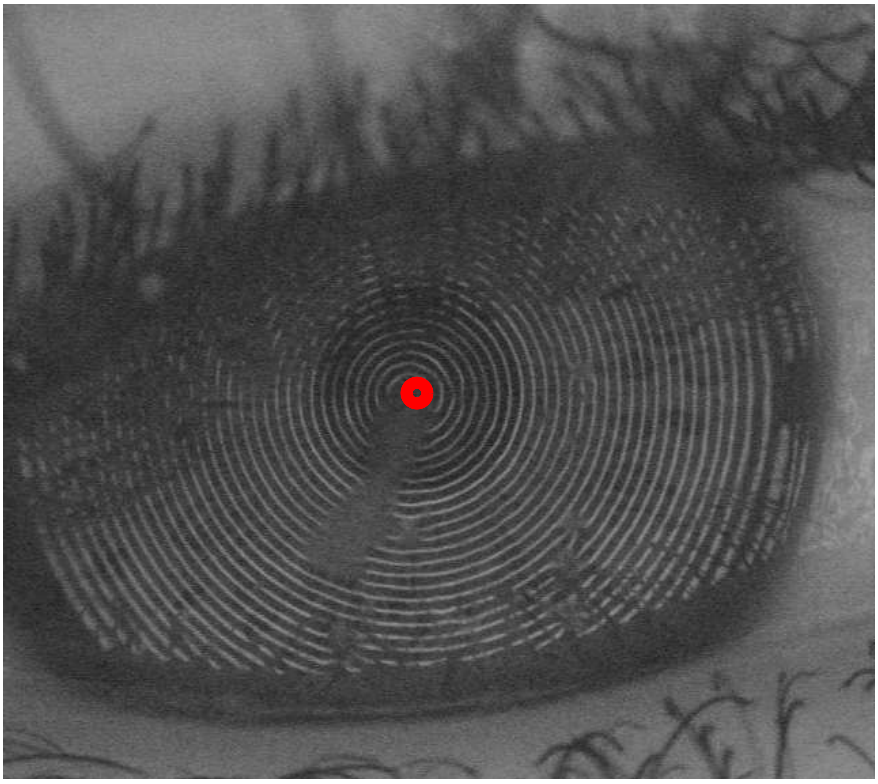
(a)

(b)

Figure 3.5. Detecting the centroid of the Placido disk pattern for a standard-acquisition case: the detected circular object and its centroid (a) and the VKS image with the its centroid (b). The large red dot is the detected centroid.



(a)



(b)

Figure 3.6. Detecting the centroid of the Placido disk pattern for a severe-mucus case: the detected arc object, the fitted circle, and its centroid (a) and the VKS image with the its centroid (b). The large red dot is the detected centroid.

highly simplifies the removal of outliers by processing in the same same direction of these quasi-lines.

Define $\delta_x = n_x - C_x$ and $\delta_y = n_y - C_y$, where n_x and n_y are the number of pixels of $I_N(x, y)$ in x and y directions, respectively. Let $\rho_{max} = \sqrt{\delta_x^2 + \delta_y^2}$ defines the maximum allowed radius for polar transformation. Let $\delta_\rho = \frac{\rho_{max}}{n_\rho - 1}$ be the radial step and $\delta_\theta = \frac{2\pi}{n_\theta}$ be the angular step, where n_ρ and n_θ are the number of pixels in the radial and angular directions of the polar image, respectively. We define two vectors, $\mathbf{P}$ and $\mathbf{\Theta}$, with the corresponding radial and angular values, respectively

$$\mathbf{P} = \left[\rho_1\ \rho_2\ \ldots\ \rho_{n_\rho}\right]^T = \left[0\ \delta_\rho\ \ldots\ (n_\rho - 1)\delta_\rho\right]^T, \tag{3.15}$$

$$\mathbf{\Theta} = \left[\theta_1\ \theta_2\ \ldots\ \theta_{n_\theta}\right]^T = \left[0\ \delta_\theta\ \ldots\ (n_\theta - 1)\delta_\theta\right]^T, \tag{3.16}$$

where $(\cdot)^T$ denotes the transpose operator.

An $n_\rho \times n_\theta$ polar grid [82] can be defined from $\mathbf{P}$ and $\mathbf{\Theta}$ according to:

$$\mathbf{G}_P = \begin{bmatrix} (\rho_1, \theta_1) & (\rho_1, \theta_2) & \cdots & (\rho_1, \theta_{n_\theta}) \\ (\rho_2, \theta_1) & (\rho_2, \theta_2) & \cdots & (\rho_2, \theta_{n_\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ (\rho_{n_\rho}, \theta_1) & (\rho_{n_\rho}, \theta_2) & \cdots & (\rho_{n_\rho}, \theta_{n_\theta}) \end{bmatrix}. \tag{3.17}$$

A Cartesian transformation is then applied to the points of $\mathbf{G_P}$ to get a Cartesian grid according to:

$$\mathbf{G}_C = \begin{bmatrix} (x_{11}, y_{11}) & (x_{12}, y_{12}) & \cdots & (x_{1n_\theta}, y_{1n_\theta}) \\ (x_{21}, y_{21}) & (x_{22}, y_{22}) & \cdots & (x_{2n_\theta}, y_{2n_\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{n_\rho 1}, y_{n_\rho 1}) & (x_{n_\rho 2}, y_{n_\rho 2}) & \cdots & (x_{n_\rho n_\theta}, y_{n_\rho n_\theta}) \end{bmatrix}, \tag{3.18}$$

where $x_{ij} = \rho_i \cos\theta_j + C_x$ and $y_{ij} = \rho_i \sin\theta_j + C_y$. We define a uniformally-spaced $n_x \times n_y$ Cartesian grid, which corresponds to the coordinates pairs of $I_N(x, y)$, according to:

$$\mathbf{G}_{UN} = \begin{bmatrix} (1, 1) & (2, 1) & \cdots & (n_x, 1) \\ (1, 2) & (2, 2) & \cdots & (n_x, 2) \\ \vdots & \vdots & \ddots & \vdots \\ (1, n_y) & (2, n_y) & \cdots & (n_x, n_y) \end{bmatrix}. \tag{3.19}$$

A 2D linear interpolation, the bilinear interpolation[2] [90], is then used to calculate the intensity values at the data points of $\mathbf{G}_C$ utilizing the intensity values of $I_N(x, y)$ at the data points of $\mathbf{G}_{UN}$. In this case, interpolation can be regarded as searching

---

[2] Bilinear interpolation is a linear interpolation in one dimension followed by a linear interpolation in the other orthogonal dimension.

procedure for the intensity value at each (x_{ij}, y_{ij}) using a look-up table, which has intensity values at fixed resolution locations determined by the coordinates pairs of $I_N(x, y)$. This yields the polar image, $I_P(x, y)$. Fig. 3.7 illustrates an example of the polar transformation of the VKS image shown in Fig. 3.1(b). The black hills are the result of interpolating at points that are outside the range of the grid. However, they have minimal effect on the quasi-lines, which are our concern.

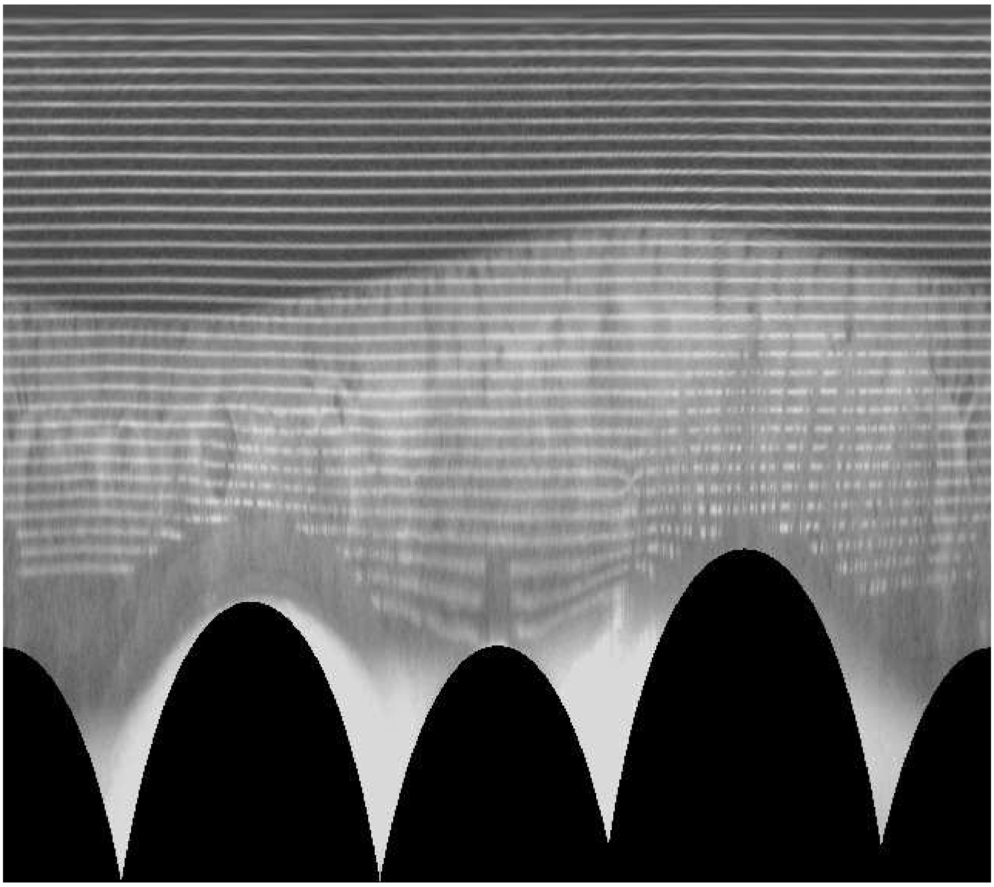


Figure 3.7. The polar transformation of the VKS image shown in Fig. 3.1(b).

3.2.3 Adaptive Spatial Filtering

An adaptive filtering approach is maintained at this stage to reduce the effect of the Gaussian-like noise and smooth the variations within the quasi-lines in the polar image version, $I_P(x, y)$, of the VKS image, while at the same time preserving their edges. This is a requirement necessary for accurate realization of the height-data estimator based on the arc-step algorithm [50]. The filter is adaptive, space-variant, in the sense that its response changes depending on the local statistical characteristics of $I_P(x, y)$.

Among many varieties, the adaptive Wiener filter is used for the problem at hands. Generally, the Wiener filter [91] works on a signal model of the form

$$I_P\left(x,y\right) = F\left(x,y\right) + N\left(x,y\right), \tag{3.20}$$

where $I_P\left(x,y\right)$ is considered as the degraded image to be filtered, $F\left(x,y\right)$ is the clean (unknown) image and $N\left(x,y\right)$ is the noise image. The clean and the noise images are assumed to be samples from linearly-independent zero-mean random processes with power spectra $P_F\left(\omega_1,\omega_2\right)$ and $P_N\left(\omega_1,\omega_2\right)$, respectively. The Wiener filter is the optimal linear minimum mean square error (MMSE) estimate of $F\left(x,y\right)$ from $I_P(x,y)$, with a frequency response $H\left(\omega_1,\omega_2\right)$ given by

$$H\left(\omega_1,\omega_2\right) = \frac{P_F\left(\omega_1,\omega_2\right)}{P_F\left(\omega_1,\omega_2\right) + P_N\left(\omega_1,\omega_2\right)}. \tag{3.21}$$

One possible modification of the Wiener filter is performed when the power spectra are estimated locally rather than estimating them globally from the whole image. Accordingly, we define a fixed neighborhood region, of size $K_1 \times K_2$, centered around each pixel in $I_P\left(x,y\right)$.

We assume that, in the model of Eq. (3.20), $N\left(x,y\right)$ has a mean value of m_N and it is white with a variance of v_N. Moreover, we assume that $F\left(x,y\right)$ is stationary in the neighborhood region $K_1 \times K_2$ with a mean value of m_F and a variance of v_F, respectively. If $m_F = m_N = 0$, the Wiener filter can be described in the neighborhood region by:

$$H\left(\omega_1,\omega_2\right) = \frac{v_F}{v_F + v_N}, \tag{3.22}$$

which can be expressed in the spatial domain as

$$h\left(x,y\right) = \frac{v_F}{v_F + v_N}\delta\left(x,y\right). \tag{3.23}$$

Here $\delta\left(x,y\right)$ is the 2D Dirac's delta function, defined as

$$\delta\left(x,y\right) = \begin{cases} 1, & x = y = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.24}$$

If m_F and m_N have non-zero values, they can be simply subtracted from $I_P\left(x,y\right)$ before filtering and then m_F can be added to the output of the Wiener filter. However, for simplicity, we assume that $m_N = 0$. Therefore, the filtered image, $I_{PF}\left(x,y\right)$, can be expressed in the local neighborhood region according to [92]:

$$\begin{aligned}
I_{PF}(x,y) &= I_P(x,y) * h(x,y) \\
&= m_F + (I_P(x,y) - m_F) * \frac{v_F}{v_F + v_N}\delta(x,y) \\
&= m_F + \frac{v_F}{v_F + v_N}\left(I_P(x,y) - m_F\right),
\end{aligned} \tag{3.25}$$

where $*$ denotes the convolution operator and $h(x,y)$ is the 2D impulse response.

The pixelwise adaptive Wiener filter, where m_F and v_F are updated pixel-by-pixel, can be expressed according to:

$$I_{PF}(x,y) = m_F(x,y) + \frac{v_F(x,y)}{v_F(x,y) + v_N}\left(I_P(x,y) - m_F(x,y)\right). \tag{3.26}$$

From Eq. (3.20) and assuming that $m_N = 0$, we can deduce

$$m_F(x,y) = m_{I_P}(x,y), \tag{3.27}$$

and

$$v_F(x,y) = v_{I_P}(x,y) - v_N. \tag{3.28}$$

The parameters $m_F(x,y)$, v_{I_P}, and v_N can be estimated as follows:

$$\hat{m}_F(x,y) = \frac{1}{K_1K_2}\sum_{n_1=x-\frac{K_1}{2}}^{x+\frac{K_1}{2}}\sum_{n_2=y-\frac{K_2}{2}}^{y+\frac{K_2}{2}} I_P(n_1,n_2), \tag{3.29}$$

$$\hat{v}_{I_P}(x,y) = \frac{1}{K_1K_2}\sum_{n_1=x-\frac{K_1}{2}}^{x+\frac{K_1}{2}}\sum_{n_2=y-\frac{K_2}{2}}^{y+\frac{K_2}{2}} \left(I_P(n_1,n_2) - \hat{m}_F(n_1,n_2)\right)^2, \tag{3.30}$$

$$\hat{v}_N = \frac{1}{n_\rho n_\theta}\sum_{n_3=1}^{n_\rho}\sum_{n_4=1}^{n_\theta}\hat{v}_{I_P}(n_3,n_4). \tag{3.31}$$

From Eqs. (3.26) and (3.28) and by including the estimates in Eqs. (3.29)–(3.31), the output of the pixelwise adaptive Wiener filter can be written as

$$I_{PF}(x,y) = \hat{m}_F(x,y) + \frac{\hat{v}_{I_P}(x,y) - \hat{v}_N}{\hat{v}_{I_P}(x,y)}\left(I_P(x,y) - \hat{m}_F(x,y)\right). \tag{3.32}$$

From Eq. (3.32), it is evident that in edge local regions, where $\hat{v}_{I_P}(x,y) \gg \hat{v}_N$, little smoothing is performed, and therefore, the edges are preserved ($I_{PF}(x,y) \approx I_P(x,y)$).

On the other hand, in non-edge local regions, where $\hat{v}_{I_P}(x,y) = \hat{v}_N$ or larger by a moderate amount, large smoothing is performed ($I_{PF}(x,y) \approx \hat{m}_F(x,y)$). It is worthwhile to observe that the term $\hat{v}_{I_P}(x,y) - \hat{v}_N$ cannot be negative because $\hat{v}_N \leq \hat{v}_{I_P}(i,j)$, as can be seen from Eq. (3.31). Fig. 3.8 illustrates the smoothed polar image version, $I_{PF}(x,y)$, of the image shown in Fig. 3.7. It is clear that the adaptive Wiener filter has smoothed the relatively stationary areas, while preserved the edges of the Placido disk pattern in the VKS image.

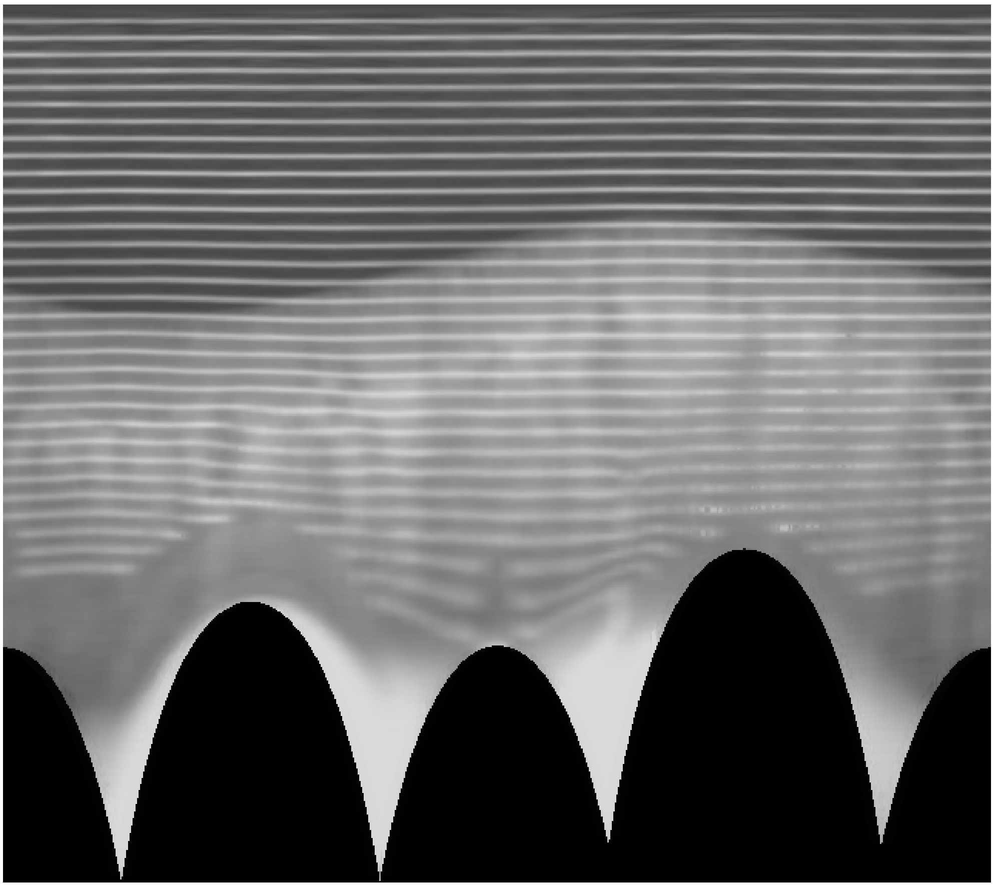


Figure 3.8. The smoothed polar image version of the image shown in Fig. 3.7

3.2.4 Morphological Closing

The Placido disk pattern of the smoothed polar VKS image, $I_{PF}(x,y)$, is still deteriorated by outliers in the form:

1. Breaks and/or gaps due to the reflections from eyelashes.
2. Fusions due to the tear-film break-up phase.

3. In some subjects, large amounts of mucus covers the eye surface.

These forms of outliers affect the reconstructed corneal topography and limit its accuracy. To suppress them, the gray-scale morphological closing operator is used. It is simply defined as a gray-scale dilation of a gray-scale image by a structuring element followed by a gray-scale erosion of the result by the same structuring element. Gray-scale dilation and erosion of $I_{PF}\left(x,y\right)$ by $SE\left(x,y\right)$ are mathematically expressed as [87]:

$$\begin{aligned}\left(I_{PF}\oplus SE\right)\left(x,y\right) &= \max\{I_{PF}(x-z_1,y-z_2)+SE(z_1,z_2)\\ &\qquad \mid (x-z_1),(y-z_2)\in D_{I_{PF}};(z_1,z_2)\in D_{SE}\}, \qquad (3.33)\end{aligned}$$

and

$$\begin{aligned}\left(I_{PF}\ominus SE\right)\left(x,y\right) &= \min\{I_{PF}(x+z_1,y+z_2)-SE(z_1,z_2)\\ &\qquad \mid (x+z_1),(y+z_2)\in D_{I_{PF}};(z_1,z_2)\in D_{SE}\}, \qquad (3.34)\end{aligned}$$

respectively. Here $D_{I_{PF}}$ and D_{SE} are the domains of $I_{PF}\left(x,y\right)$ and $SE\left(x,y\right)$, respectively. The gray-scale closing is then given by

$$I_{PF}\left(x,y\right)\bullet SE\left(x,y\right)=\left(I_{PF}\left(x,y\right)\oplus SE\left(x,y\right)\right)\ominus SE\left(x,y\right). \qquad (3.35)$$

The structuring element $SE\left(x,y\right)$ is an important parameter in our application as its orientation is selected to be in the same direction of the Placido disk pattern of the polar VKS image (i.e., horizontal direction) and its size can be tuned to fit the size of the outliers.

In the closing operation, dilation is used to enlarge the original bright areas in $I_{PF}\left(x,y\right)$ by outputting the maximum gray-level value of the region spanned by the translated $SE\left(x,y\right)$, at each pixel location. At this stage, gaps and breaks are bridged and corrected. However, the Placido disk pattern is thickened due to using the maximum operator. To suppress these artefacts, erosion is maintained by outputting the minimum gray-level value of the region spanned by the translated $SE\left(x,y\right)$, at each pixel location in the dilated $I_{PF}\left(x,y\right)$. Accordingly, any irregularities smaller than the size of $SE\left(x,y\right)$ are smoothed out and the effect of thickening the Placido disk pattern is eliminated.

The important fact is that the closing operation has left the outliers-free details without any geometrical distortions as can be seen by comparing Fig. 3.7 with its smoothed and closed version shown in Fig. 3.9.

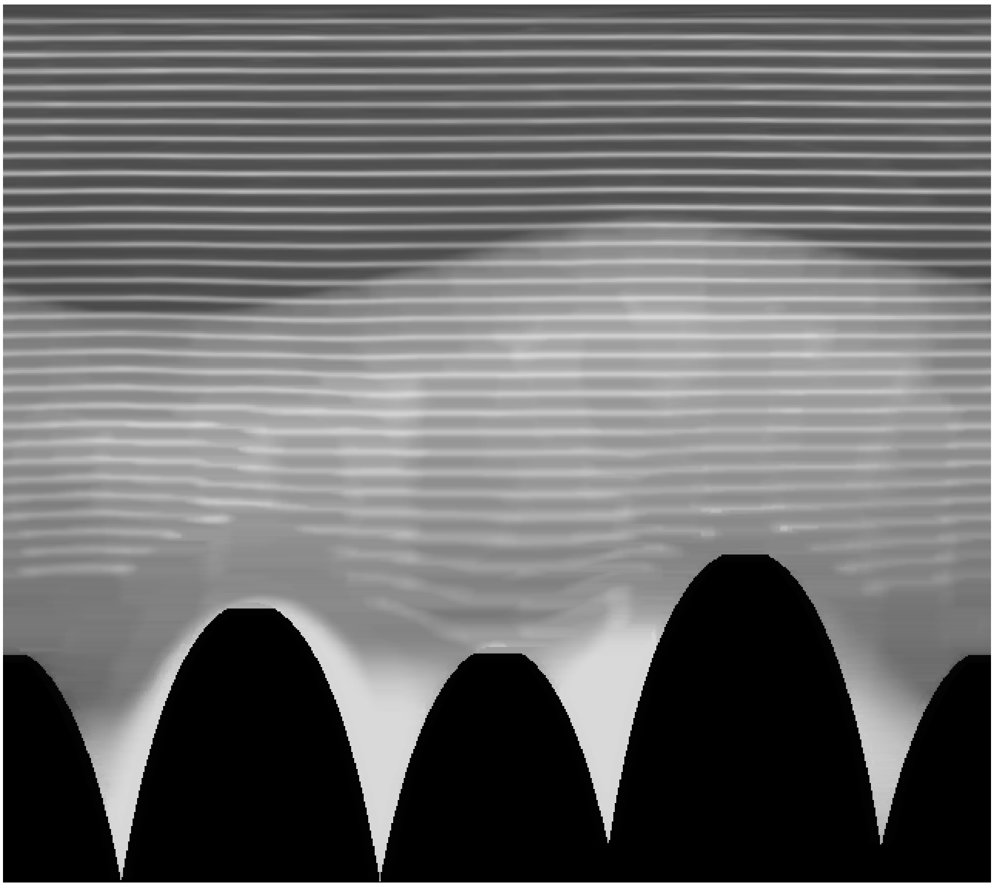


Figure 3.9. The smoothed and closed polar image version of the image shown in Fig. 3.7

3.2.5 Cartesian Transformation

Cartesian Transformation is applied to get the final processed VKS image back to its original domain. From Eqs. (3.18) and (3.19), $\mathbf{G}_C$ and $\mathbf{G}_{UN}$ can be used for this purpose. Note that the interpolation should be maintained in order to result in an intensity image at uniformally-spaced grid coordinates, $\mathbf{G}_{UN}$, from a nonuniformally-spaced grid coordinates, $\mathbf{G}_C$. First, a surface is fitted to the intensity values of the points of $\mathbf{G}_C$. Then, the surface is interpolated, using triangle-based linear interpolation, at the points of $\mathbf{G}_{UN}$.

In more details, this procedure depends on Delaunay triangulation [93], in which each point of $\mathbf{G}_C$ is connected to its neighbors by lines in the form of triangles. These triangles are not random and they obey the restriction that no point in $\mathbf{G}_C$ is inside the circumcircle of any generated triangle. Delaunay triangulation can be reduced to finding the convex hull of the data points of $\mathbf{G}_C$. Convex hull of a set of points is defined as the smallest convex set [94] that contains the points. A set C is convex if for any $p_1, \ldots, p_r$ in C and non-negative numbers $\lambda_1, \ldots, \lambda_r$ such that $\sum_{i=1}^{r} \lambda_i = 1$, then the vector $\sum_{i=1}^{r} \lambda_i p_i$ is in C. A well-known algorithm for computing the convex hull, Qhull, is available in [95]. It is clear that this triangulation strategy is used for building a mesh from the points of $\mathbf{G}_C$. Then, a search algorithm is maintained in order to assign a triangle (the nearest triangle) for each point in $\mathbf{G}_{UN}$. Finally, barycentric coordinates on triangles [96] are used to perform linear interpolation to calculate the intensity value for each point in $\mathbf{G}_{UN}$ from the known intensity values of the vertices of the triangle that contains it. This is performed by assuming that p is a point of $\mathbf{G}_{UN}$ contained in a triangle with vertices v_1, v_2 and v_3. Using barycentric coordinates, p can be written according to

$$p = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3, \tag{3.36}$$

where $\lambda_1 + \lambda_3 + \lambda_3 = 1$, and p and each vertex have three components, (x, y, z), where z denotes the intensity value. The intensity value, z, of the unknown point, p, can be calculated by solving the linear set of equations in Eq. (3.36). Fig. 3.10 illustrates the Cartesian processed VKS image, $\hat{\mathbf{I}}_N(x, y)$, whose original was shown in Fig. 3.1(b).

3.3 Performance Evaluation

The corneal height-data from the videokeratoscope, $S(\rho, \theta)$, are usually modeled by a finite set of orthogonalized Zernike polynomials[3] [4]. This can be expressed, according to the single-indexed notation [73], as follows

$$S(\rho, \theta) = \sum_{k=1}^{K} a_k Z_k(\rho, \theta) + \varepsilon(\rho, \theta), \tag{3.37}$$

where $Z_k(\rho, \theta)$ is the single indexed k-th Zernike polynomial with a coefficient a_k and $\varepsilon(\rho, \theta)$ represents the measurement and modeling error. A basic question is how many Zernike terms should be used. Using a bootstrap approach [10, 12, 13][4], it has been found that a relatively low radial order of the Zernike polynomial expansion is sufficient for modeling healthy corneas [6, 7]. There, the optimal number of Zernike terms was decided by choosing the model-order that minimizes the bootstrap mean-square error. Later, it has been found that this procedure may underestimate the clinically expected model order due to the assumption that the $\varepsilon(\rho, \theta)$ is independent and identically distributed (i.i.d) across the whole corneal surface. In [8], the bootstrap procedure

---

[3] Details of using Zernike polynomials in modeling of corneal surfaces are found in Chapter 2.

[4] Full details of using resampling-based techniques (i.e., bootstrap and hook-and-loop (HL)) for model-selection are provided in Appendix A.

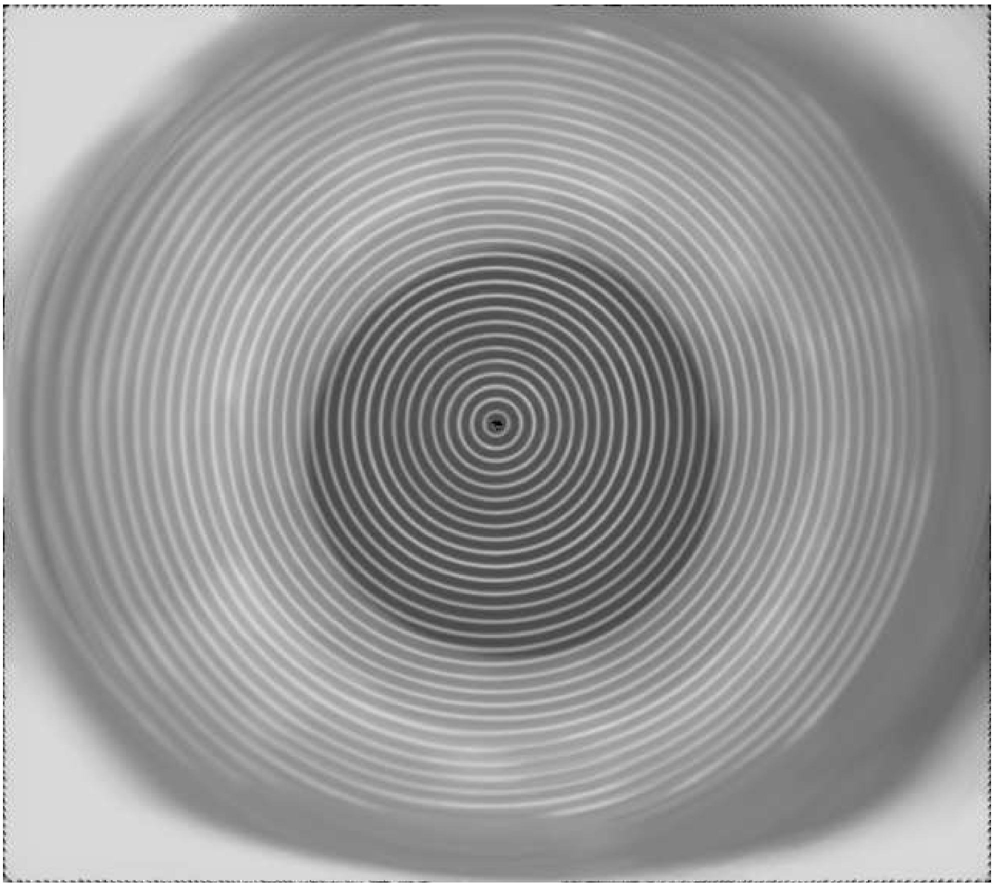


Figure 3.10. The Cartesian processed VKS image of the image shown in Fig. 3.1(b).

has been refined by incorporating the knowledge of the spatially non-uniformity of the measurement noise in the resampling procedure. This was achieved by introducing dependency of $\varepsilon(\rho, \theta)$ on ρ. This means that for a given ρ, $\varepsilon(\rho, \theta)$ is assumed as an i.i.d random variable. However, this procedure is numerically much more complex than its predecessor and hence, has a very limited practical applicability. One main problem in using bootstrap for model-order selection is the necessity of selecting a suitable scaling parameter, namely m, for the detrended residuals in the resampling procedure [97]. An alternative to scaling is based on subsampling [98]. Nevertheless, estimating the subsample length is a problem particularly when the sample length is small.

A new resampling scheme, called the "Hook-and-Loop" (HL), has been recently proposed for avoiding scaling problems [21]. In the proposed HL resampling scheme, it was suggested that the concept of the non-parametric bootstrap of sampling from the empirical distribution function is taken rather literally. The idea behind HL is to sort the residuals to match them to the strength of the original signal in order to eliminate the need for scaling the residuals. Moreover, the HL has been used for estimating the optimal number of Zernike terms when fitted to different real corneal height-data [20]. HL-based model-order selection has resulted in a number of Zernike terms that, in some cases, agree with clinical expectations and is very similar to that obtained with the refined bootstrap method presented in [8] but with a computational cost similar to that of the standard bootstrap procedure presented in [6].

Accordingly, we statistically assess the performance of our proposed adaptive image processing procedure using prediction errors that depend on the HL resampling plane.

3.4 Experimental Results with Discussion

To elucidate the power of the proposed image processing procedure, we applied it to the original videokeratoscopic measurements of test surfaces and a range of real corneas. Here, we present the results of six representative cases including

1. An 8 mm PMMA sphere with a fingerprint, Fig. 3.1(a)
2. A standard case with minimal interference, Fig. 3.1(b)
3. Non-standard case A, with long eye-lashes, Fig. 3.1(c)
4. Non-standard case B, with significant tear-film break-up, Fig. 3.1(d)
5. Non-standard case C, with long eye-lashes and mucus, Fig. 3.1(e)
6. Non-standard case D, with severe mucus, Fig. 3.1(f)

As shown in Fig. 3.2 and described in section 3.2, we process the raw Placido disk image using the proposed procedure and feed it back to the instrument's corneal topography estimation algorithm.

As previously mentioned, precise centroid detection is essential for the subsequent steps of the procedure. If a random centroid is used in polar image transformation, the Placido ring pattern (or the quasi-straight lines in the polar version of the VKS image) would have been distorted. Therefore, subsequent processing would result in rings fusion, probably complete washing out of the image, and consequently incorrect corneal height-data. We tested the centroid detection procedure on a range of real non-rotationally symmetric corneal surfaces (some with high astigmatism and some keratoconic) and bench-marked them against the instrument's own estimator that is provided in Medmont Export File (MXF). We found a very close (sub-pixel) agreement between the two detection methods. Moreover, even for highly-occluded central parts of the corneal surfaces (e.g., Fig. 3.1(f)), the centroid detection algorithm was able to detect the centroid within the innermost ring. We acknowledge that the proposed centroid detection algorithm can be further improved. However, centroid estimation was not the primary goal of this work and its only purpose was the self-sufficiency and completeness so we do not need to rely on information from the MXF file.

In the experiments, we used the following settings:

1. For the centroid detection procedure, there was an extensive use of morphological operators. A 3×3 structuring element of a diamond shape was used for morphological reconstruction in the edge-contours subimage, $J_{BW}(i, j)$. A flat structuring element of size 5×5 was used in the opening operation for eliminating trivial edge contours. In case of detecting a circular object, we have used the constraint that it should contain at least 100 foreground pixels to be considered as a relatively-large object. For detected thinned arc objects, at least 50 points have been used for fitting the circle to avoid instabilities and to suppress trivial and broken arcs.
2. The resolution of the polar grid, $\mathbf{G}_P$, was chosen to be the same as for the Cartesian grid, $\mathbf{G}_C$. Hence, $n_\rho = n_y$ and $n_\theta = n_x$.
3. The neighborhood region in the adaptive pixel-wise Wiener filter was 1×15 to maintain smoothing in the horizontal direction, which is the main orientation of the Placido disk pattern in the polar version of the VKS image.
4. The size of the structuring element in the gray-scale morphological closing operation was 1×31. Again, the main orientation of the Placido disk pattern has controlled the direction of the structuring element.

To evaluate the accuracy of the adaptive image processing procedure, we took the 8 mm PMMA sphere that is usually used for calibrating corneal topography instruments. The original accuracy of such a calibrating sphere is at least an order higher (about 100 nm) than the expected videokeratoscopic measurement error and is determined by a surface profiling instrument. To introduce some “noise”, a fingerprint was left on the surface, as can be seen in the lower half part of Fig. 3.1(a). Figs. 3.11 and 3.12 illustrate the original videokeratoscopic image of the considered spherical surface and the one that was processed with our technique, along with their corresponding height-data estimated by the instrument’s own algorithm. The output data of the majority of Placido disk videokeratoscopes consists of a point cloud of corneal height-data sampled at a range of radial and azimuthal points. The number of radial samples depends on the number of Placido disk rings (in our case up to 32) while the number of azimuthal samples is in the range of several hundreds (in our case 300). In the case of the calibrating sphere, the parametric model was known and a four-parameter sphere (x_0, y_0, z_0, R) was fitted to both sets of corneal topography estimates from the original and processed images. The image processing procedure showed significant performance in this case, as the RMS error (height residuals) was reduced from 0.25 μm (for the original image) to 0.18 μm. Also, the outliers in the original image, that were due to the fingerprint left on surface, have been corrected.

In the case of real corneas, determining the accuracy of a videokeratoscopic measurement is not so straightforward. One way of assessing the performance of the method is to use the same parametric model and compare its fit to the height-data estimated from the original image to that from the processed image. To do this, series of Zernike polynomials of radial orders ranging from 6 to 12 (i.e., 21 to 91 coefficients) have been fitted to an 8 mm diameter of the corneal surface. Table 3.1 shows the values of the model fit RMS errors for height-data estimated from the six original images shown in Fig. 3.1, including the calibrating sphere, and estimated from the processed images using our proposed technique (numbers shown in bold font). The RMS error (in μm) for k Zernike coefficients was calculated according to

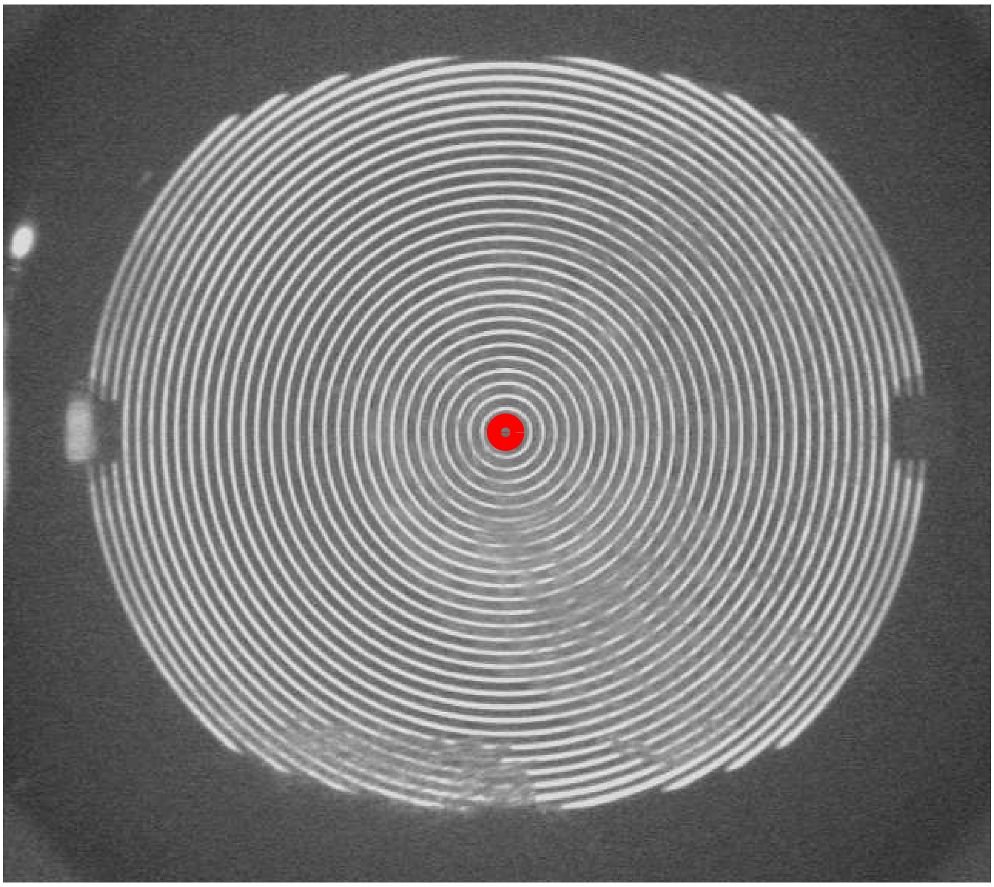
$$\mathrm{RMSE}_k = 1000\sqrt{\frac{1}{N}\|\boldsymbol{S} - \boldsymbol{Z}_k\hat{\boldsymbol{a}}_k\|^2}, \tag{3.38}$$

where N is the number of topographical points of the corneal surface, $\boldsymbol{S}$ is an $N \times 1$ vector of topographical points, $\boldsymbol{Z}$ is an $N \times K$ matrix of discrete, orthogonal Zernike polynomials, and $\hat{\boldsymbol{a}}_k$ is the least-squares estimator [71] of the parameter vector $\boldsymbol{a}$ conditioned on k. Reduction in the model fit RMS error was registered for all radial orders and all cases, including that of a standard acquisition.

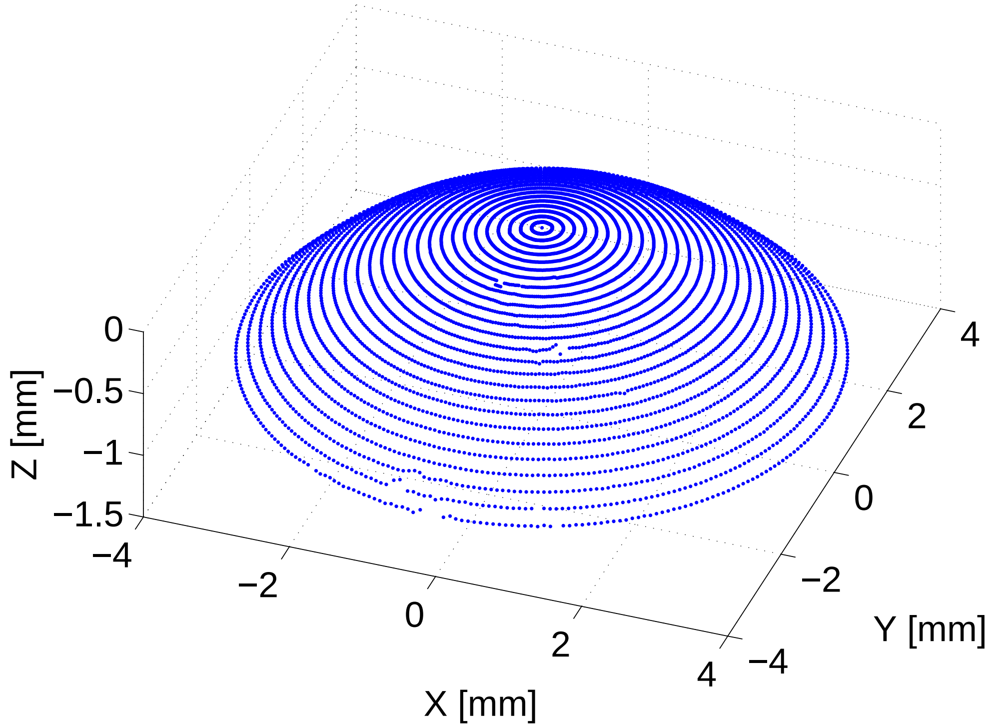
An attractive property of our procedure is increasing the corneal surface coverage. To measure this parameter, we calculated the relative increment in the number of the topographical points retrieved by the videokeratoscope after applying the procedure. Denote the number of retrieved topographical points for the original and the processed VKS images as S_O and S_P, respectively. Then, the relative increment of the corneal surface coverage is calculated as

$$S_{\mathrm{inc.}} = \frac{S_P - S_O}{S_O} \times 100\%. \tag{3.39}$$

The results are shown in the last column of Table 3.1.

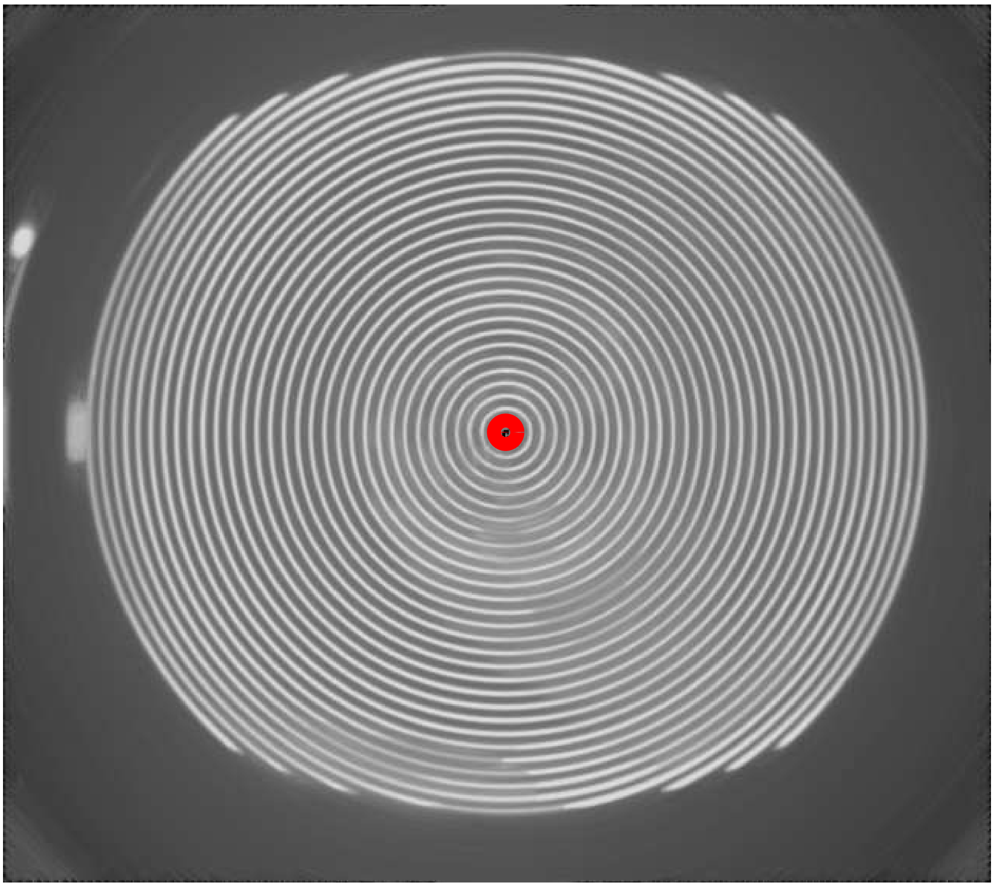


(a)

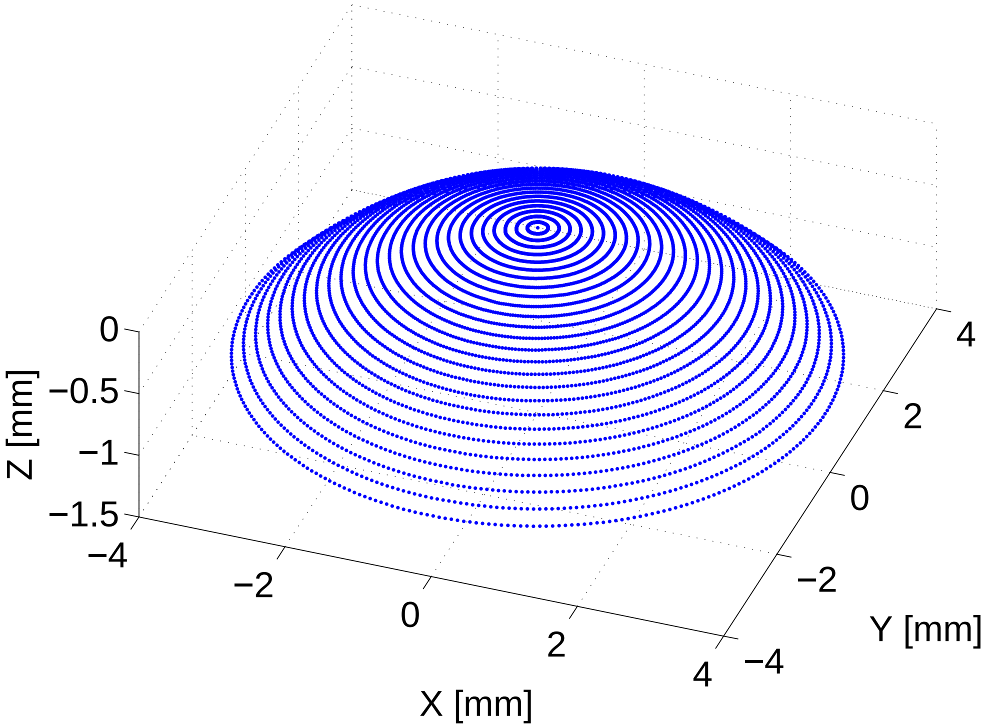


(b)

Figure 3.11. An 8 mm PMMA sphere with a fingerprint: (a) the raw VKS image, and (b) its corneal topography.



(a)



(b)

Figure 3.12. An 8 mm PMMA sphere with a fingerprint: (a) the processed VKS image, and (b) its corneal topography.

Table 3.1. The height residuals RMS errors (in μm) of fitting an 8 mm diameter of the corneal surface height-data with sets of Zernike polynomials of radial orders: 6, 8, 10 and 12 for the original image (Org.) and the processed image (Proc.) as well as the increase in the surface coverage area (last column).

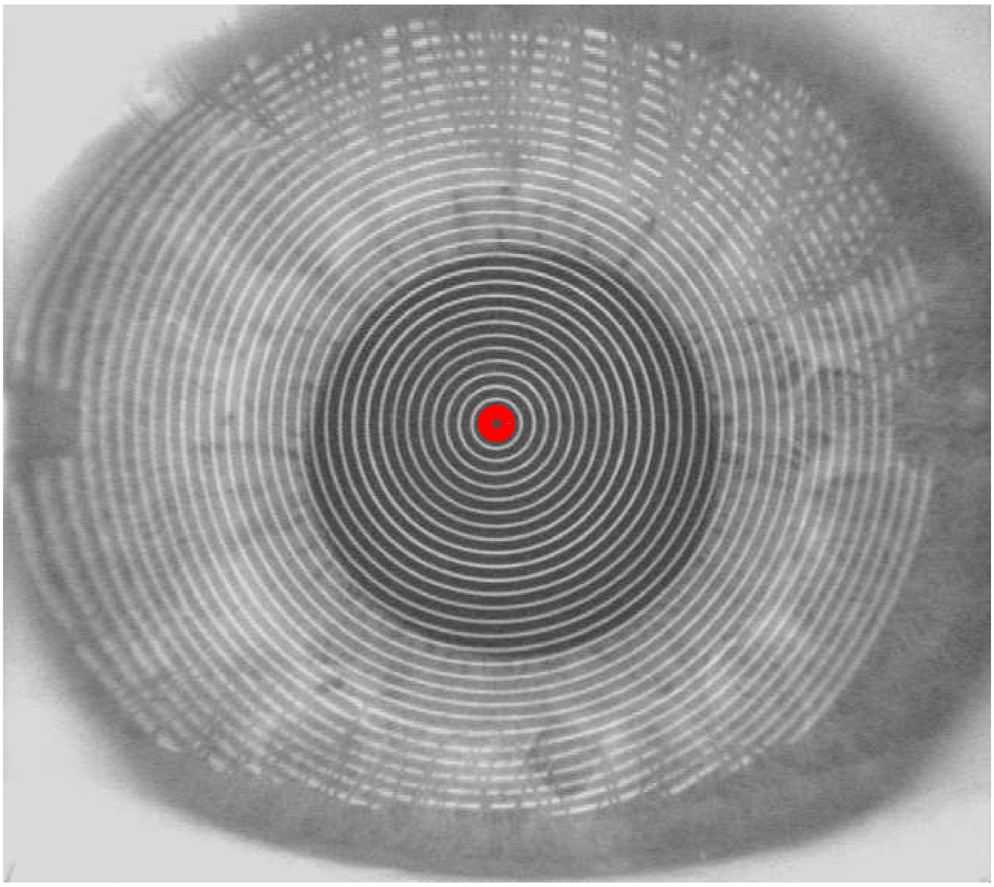
Image Example	Zernike polynomials radial order								Increase in Surface Coverage
	6		8		10		12		
	Org.	Proc.	Org.	Proc.	Org.	Proc.	Org.	Proc.	
Spherical Model	0.21	**0.08**	0.20	**0.07**	0.20	**0.06**	0.19	**0.05**	12.18%
Standard Acquisition	0.20	**0.15**	0.17	**0.12**	0.14	**0.08**	0.13	**0.07**	1.54%
Long Eye-lashes	0.52	**0.51**	0.49	**0.43**	0.47	**0.37**	0.46	**0.33**	7.26%
Tear-film Break-ups	1.60	**1.21**	1.51	**1.13**	1.42	**1.01**	1.34	**0.90**	6.24%
Mucus	0.96	**0.67**	0.88	**0.59**	0.78	**0.52**	0.70	**0.45**	9.35%
Severe Mucus	2.69	**0.99**	2.44	**0.89**	2.22	**0.80**	2.03	**0.74**	17.10%

Since the true corneal topography is unknown, we utilize the HL resampling plane for selecting the optimal number of Zernike terms[5], $\hat{k}_o$, for both the original and the processed data and compare their bootstrap RMS errors, $\hat{\Gamma}_{HL}$. This is shown in Table 3.2 for the six considered original data sets and the corresponding corneal height estimates derived from the processed images (numbers shown in brackets with bold font) for a range of corneal diameters. We observe that in all considered cases and all corneal diameters the estimated bootstrap root mean square error, $\hat{\Gamma}_{HL}$, was smaller for the data from the processed image (up to 63%), even in the case where the estimated optimal model-order, $\hat{k}_o$, was smaller than that of the original data (see the case of the severe mucus for an 8 mm corneal diameter). Also, in six cases the estimated model-order for the original data was too low (i.e., $\hat{k}_o < 11$) indicating poor data that are clinically unaccepted. However, in two out of these six cases (tear-film break-up case for an 8 mm corneal diameter and the mucus case for a 4 mm corneal diameter) the proposed filtering technique resulted in better model-order estimates.

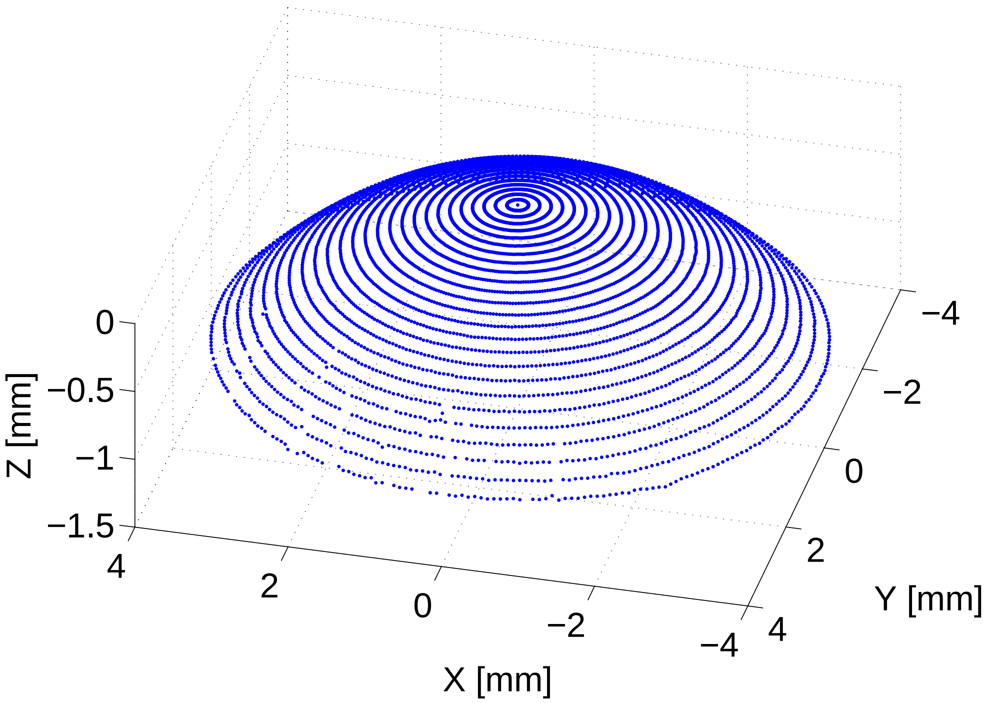
In the following, let us look more closely at the five real corneas listed earlier. Figs. 3.13 and 3.14 show the results of the standard acquisition case where minimum interference from the reflection of eyelashes and tear-film break-ups are found. However, there is still an advantage by using our image processing procedure for such an VKS image in terms of correcting for the interferences, resulting in an (slight) increment of surface coverage, and reducing the RMS error after fitting the corneal height-data with parametric models of different orders.

---

[5] Model-order selection for corneal surfaces is detailed in Chapter 4.

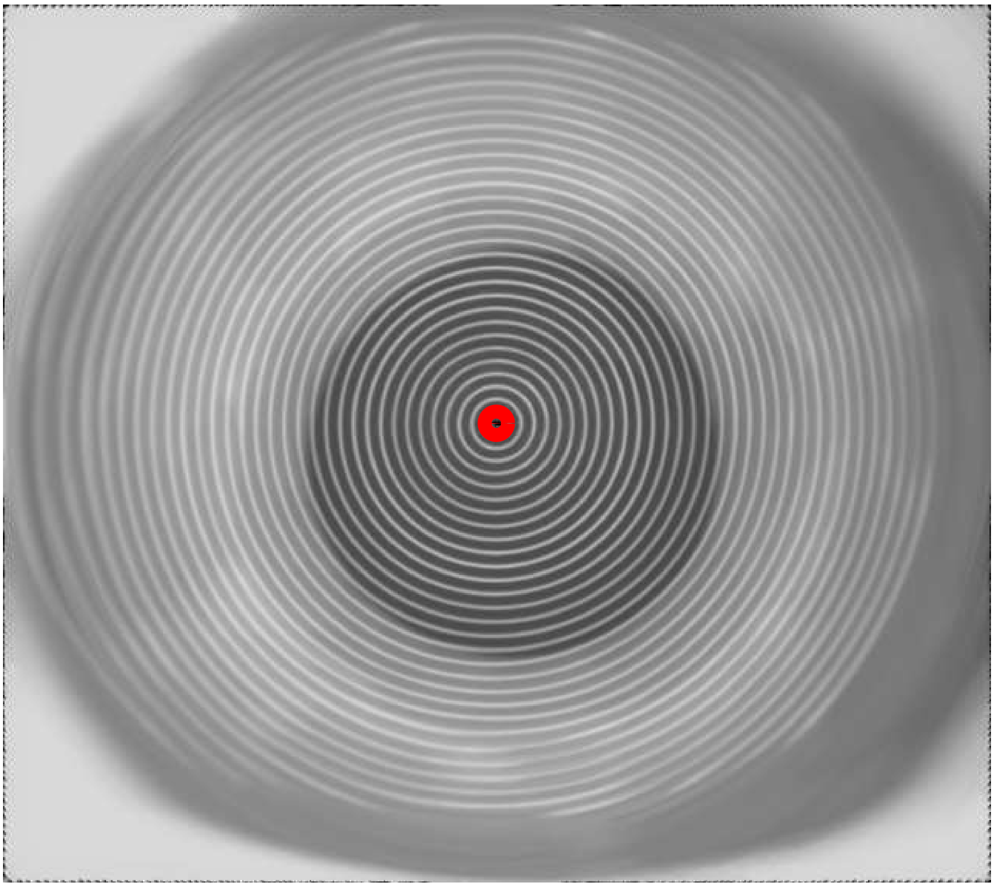


(a)

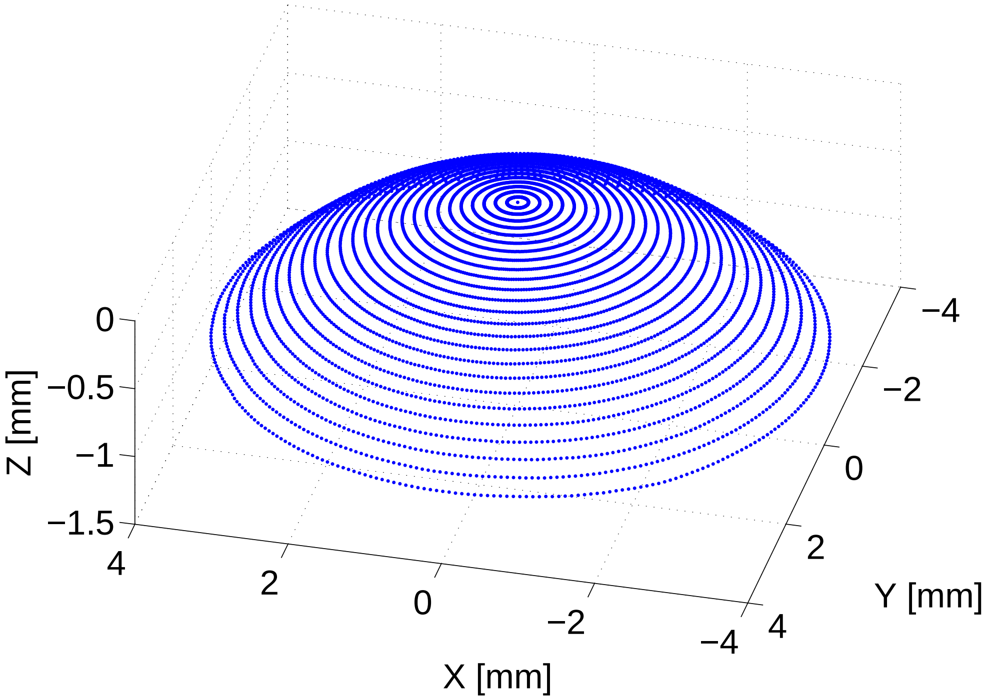


(b)

Figure 3.13. A standard acquisition case: (a) the raw VKS image, and (b) its corneal topography.



(a)



(b)

Figure 3.14. A standard acquisition case: (a) the processed VKS image, and (b) its corneal topography.

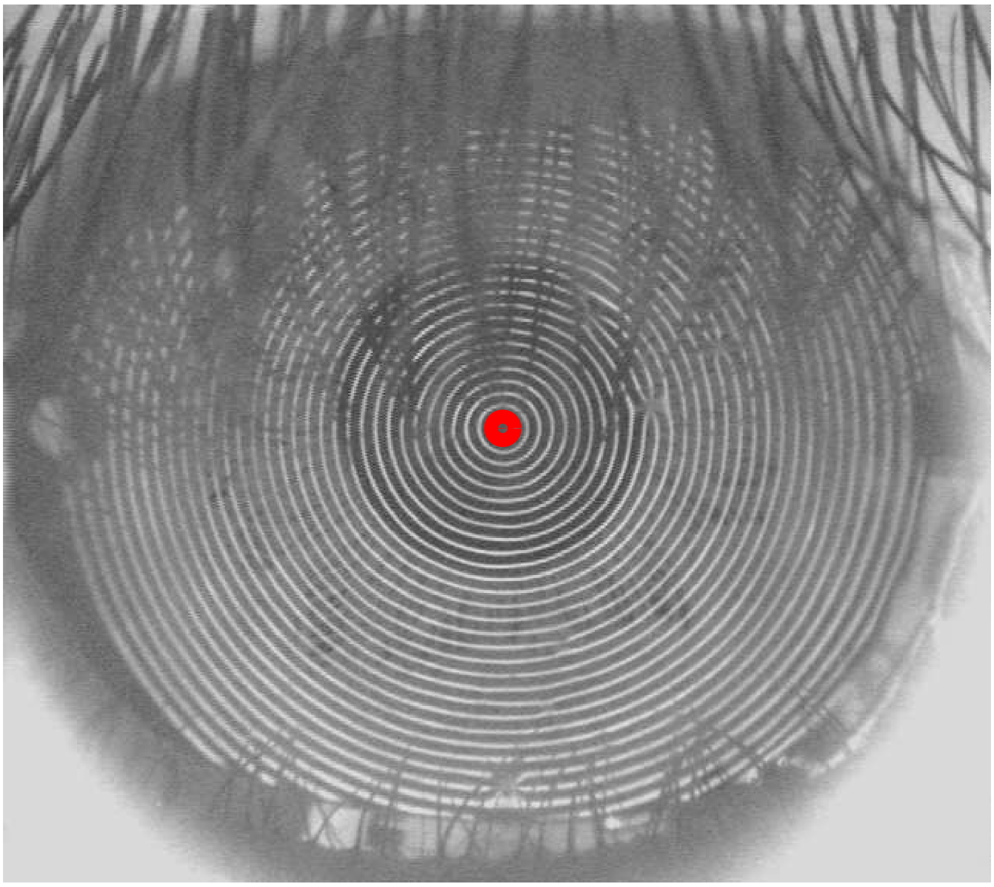
Table 3.2. The estimated optimal order, $\hat{k}_o$, of the Zernike polynomial expansion selected by the HL and the bootstrap mean square-error, $(\hat{\Gamma}_{HL})$, (in μm) of fitting a 4-, 6- and 8 mm diameter of the corneal surface height-data from the original image and the processed one (bold font in brackets).

Image Example	Corneal diameter (mm)					
	4		6		8	
	$\hat{k}_o$	$\hat{\Gamma}_{HL}$	$\hat{k}_o$	$\hat{\Gamma}_{HL}$	$\hat{k}_o$	$\hat{\Gamma}_{HL}$
Spherical Model	11(**11**)	2.35(**1.06**)	11(**12**)	2.86(**1.12**)	11(**11**)	2.97(**1.19**)
Standard Acquisition	11(**12**)	1.21(**0.72**)	12(**16**)	1.59(**1.02**)	17(**17**)	2.23(**1.42**)
Long Eye-lashes	11(**11**)	3.48(**1.98**)	12(**12**)	6.13(**3.32**)	11(**11**)	6.82(**4.41**)
Tear-film Break-ups	4(**5**)	16.50(**13.87**)	4(**4**)	20.64(**16.89**)	4(**12**)	21.56(**16.64**)
Mucus	4(**11**)	5.99(**3.38**)	11(**11**)	9.29(**5.79**)	11(**11**)	12.03(**7.99**)
Severe Mucus	4(**4**)	26.49(**13.01**)	4(**4**)	27.01(**12.93**)	44(**36**)	34.57(**12.83**)

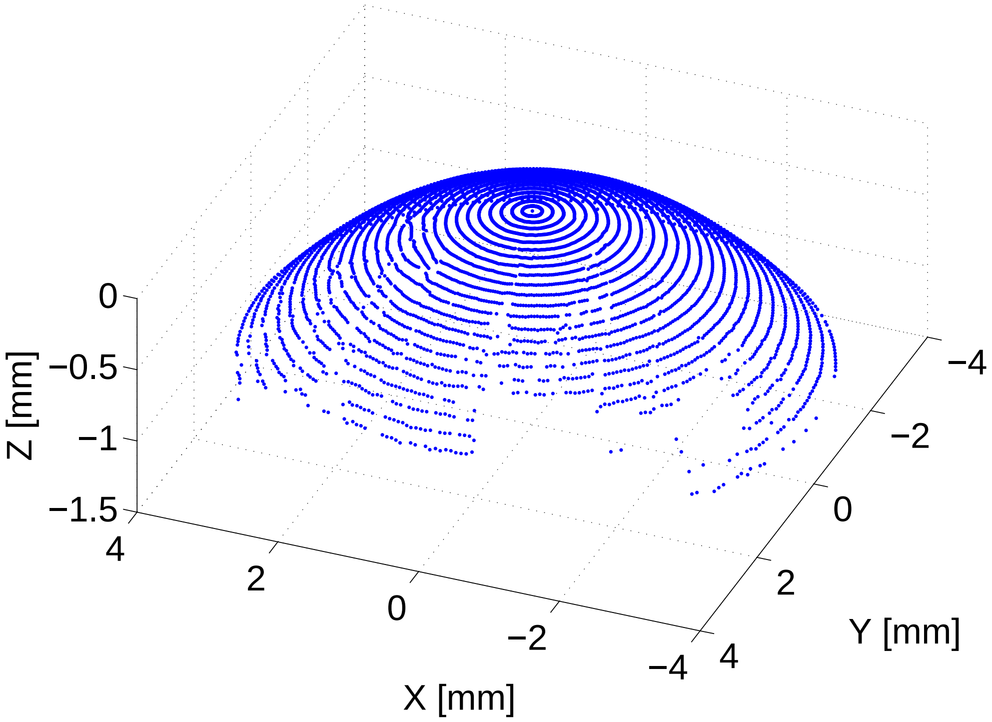
Figs. 3.15 and 3.16 illustrate the results for the non-standard case A where reflections from long eyelashes have obscured a large area in the upper half of the corneal surface. It is clear that after processing the VKS image most of the missing observations have been bridged, subsequently leading to reduction in the RMS of the parametric model fit to the height-data. However, a small number of peripheral artefacts has been introduced by the image processing procedure. These artefacts appear only at the image boundary due to the intersection between the long eyelashes and the sclera. However, they are only of minor significance because in the majority of corneal topography measurements focus is given to the central not the peripheral data. Subsequently, such artefacts can be interpreted as rings by the instrument's own algorithm.

Figs. 3.17 and 3.18 illustrate the results for the non-standard case B where significant tear-film break-up is encountered. The image processing procedure has corrected large areas of break-ups, even in parts around the centroid where the rings have shown a significant level of fusion. However, in this particular example, the videokeratoscope has failed to detect part of one ring around the centroid (seen as a small gap in the height-data of the processed image). This is despite the fact that the ring shows good contrast in the processed image. Fitting Zernike polynomials has shown a significant reduction trend in the RMS error for the corneal surface of the processed image.

Figs. 3.19 and 3.20 illustrate the results for the non-standard case C where significant reflections from eyelashes and tear-film break-ups are found in the upper half of the corneal surface, and rings washing out (due to some mucus) are encountered in the lower part of the corneal surface. The proposed image processing procedure has corrected

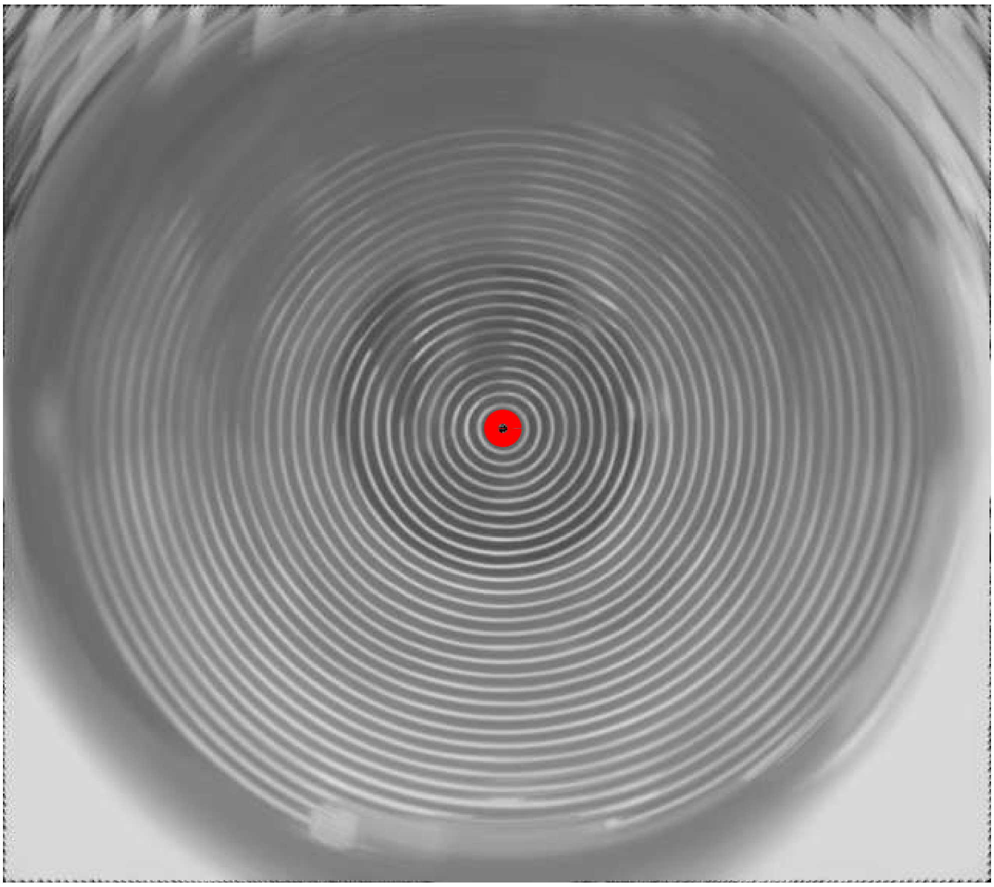


(a)

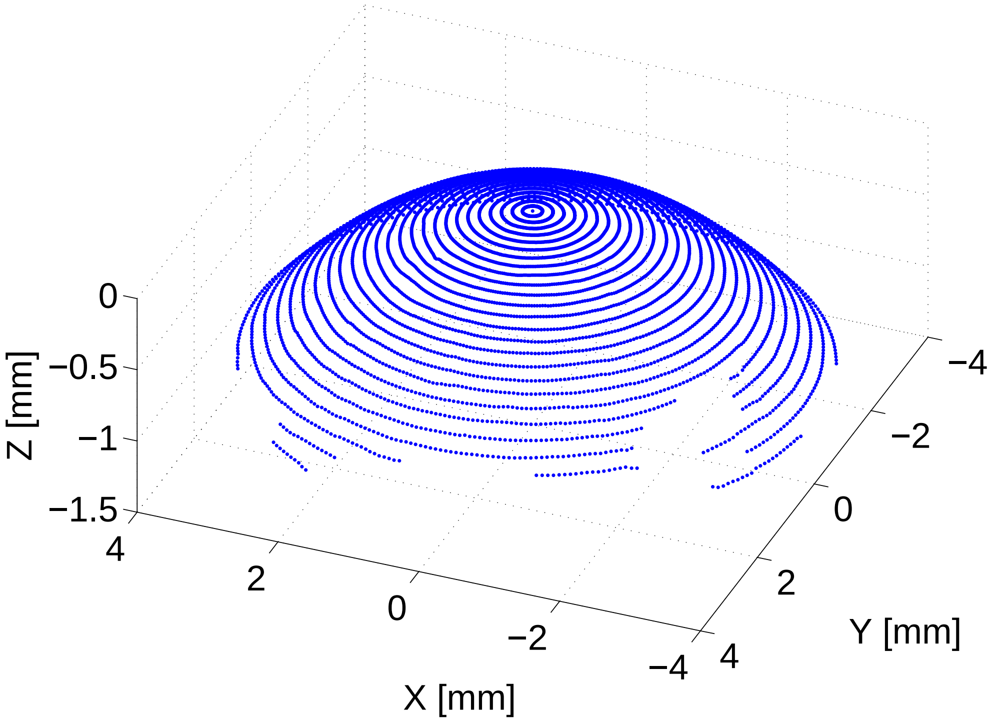


(b)

Figure 3.15. A non-standard acquisition case A (eyelash shadows from the upper and lower eyelids): (a) the raw VKS image, and (b) its corneal topography.

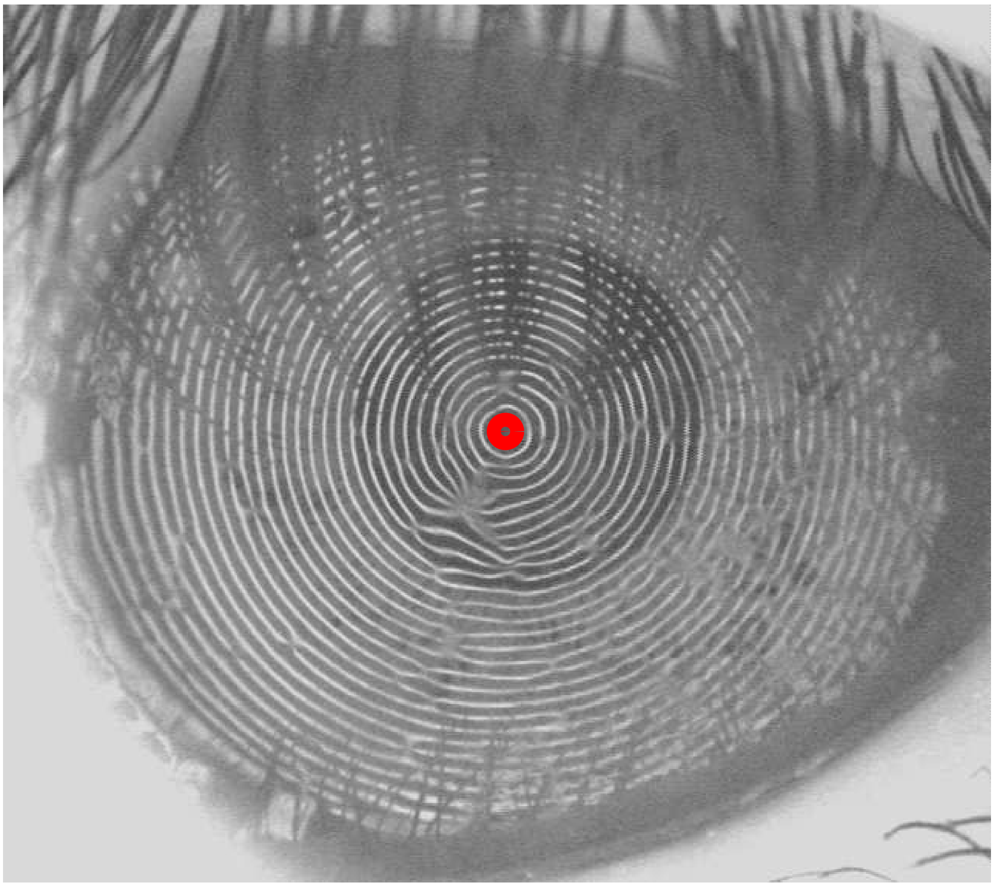


(a)

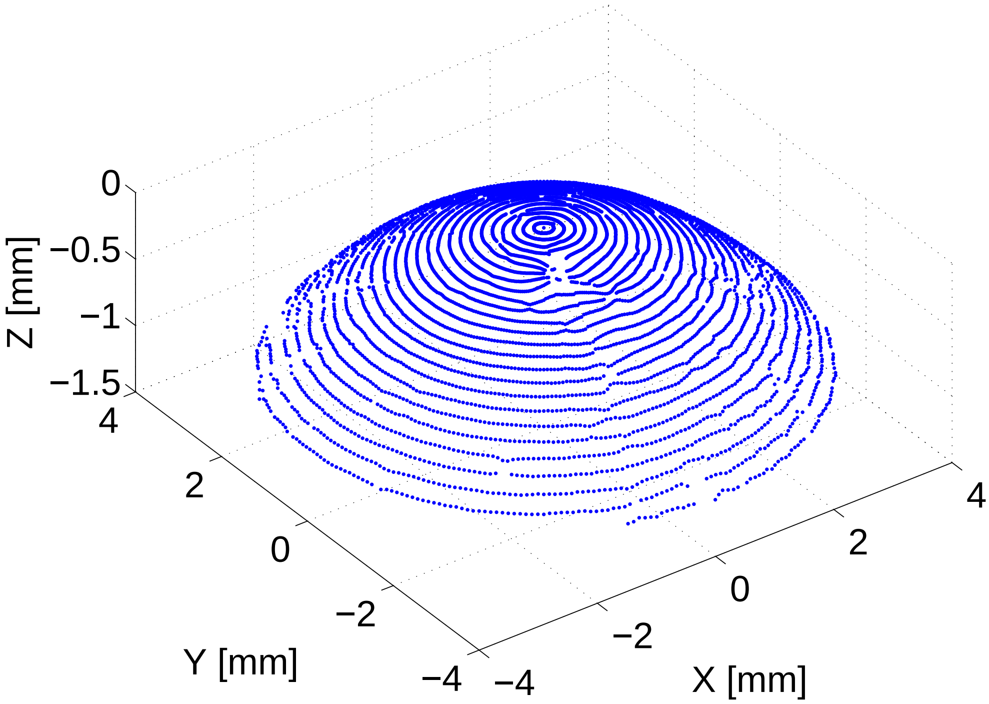


(b)

Figure 3.16. A non-standard acquisition case A (eyelash shadows from the upper and lower eyelids): (a) the processed VKS image, and (b) its corneal topography.

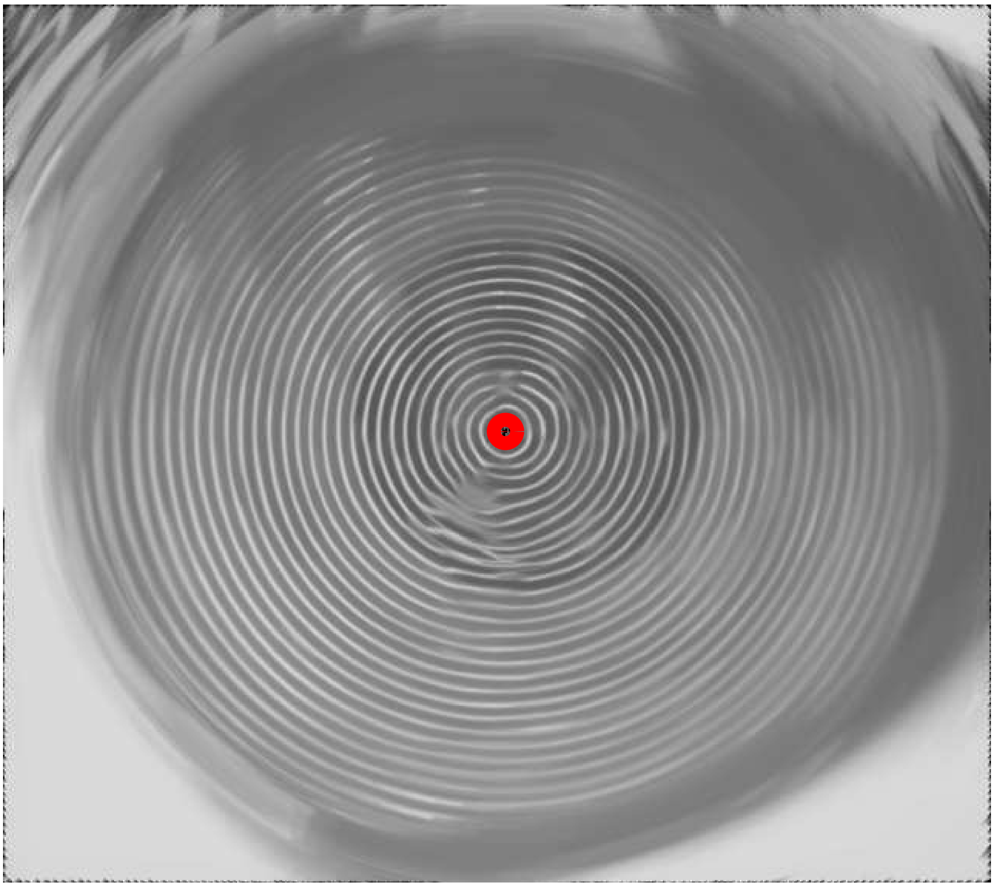


(a)

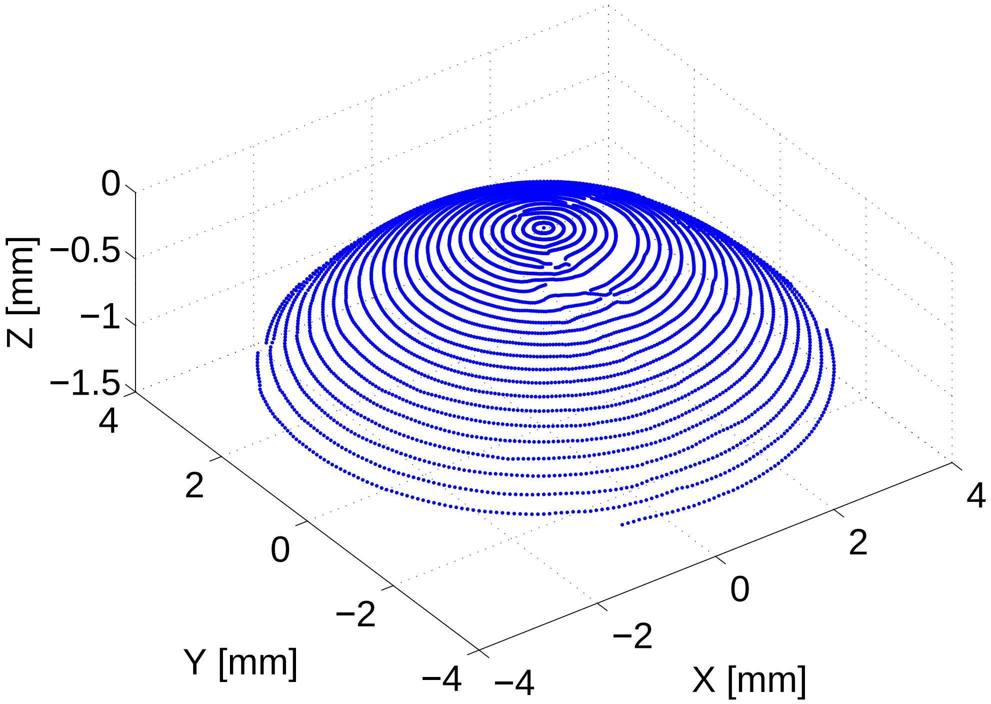


(b)

Figure 3.17. A non-standard acquisition case B (significant tear-film break-up): (a) the raw VKS image, and (b) its corneal topography.



(a)



(b)

Figure 3.18. A non-standard acquisition case B (significant tear-film break-up): (a) the processed VKS image, and (b) its corneal topography.

almost all the reflections and recovered the blurred area of the cornea. Moreover, RMS error reduction was observed when Zernike polynomials were fitted to both the original corneal surface and the reconstructed one from the processed image.

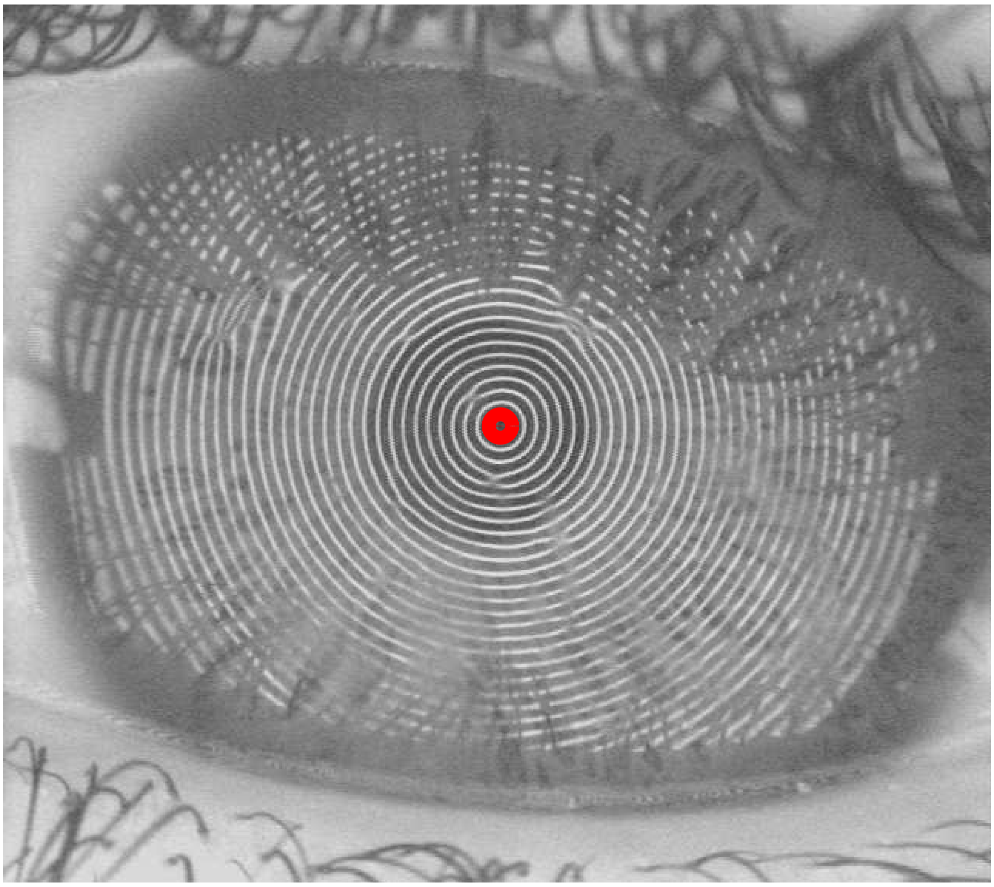
Figs. 3.21 and 3.22 illustrate the results for non-standard case D where a large amount of mucus covers the central part of the cornea. This leads to a significant amount of missing observations as can be seen from the original image and its corneal topography. Our image processing procedure has recovered a significant portion of the occluded parts, as can be observed in the processed image and its corresponding corneal topography.

3.5 Summary

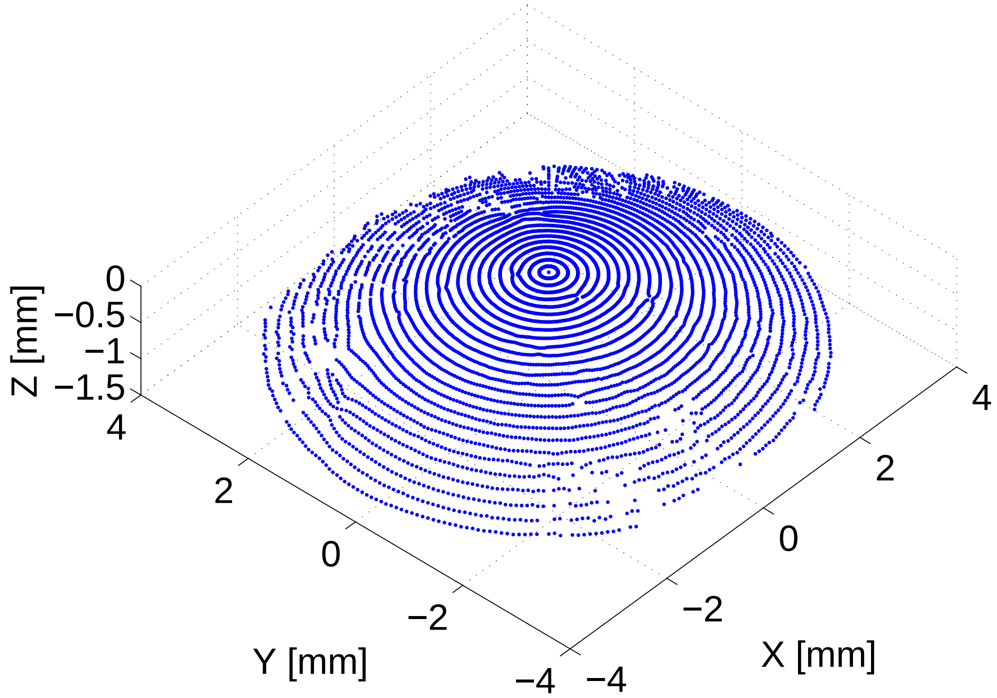
We have proposed an approach for enhancing the operating range in videokeratoscopy. This was achieved by image-processing the original Placido disk rings image before estimating the corneal topography. The approach results in a larger coverage corneal surface area by removing different types of missing observations, which may occur during non-ideal image acquisition scenarios. We showed, using test surfaces, that the proposed technique leads to more accurate corneal topography estimation and additionally results in a greater coverage area.

The procedure is based on processing the polar-transformed version of the Placido disk rings image with a combination of pixel-wise adaptive filter and morphological closing operation. Experimental results have shown that the procedure works well in all considered cases. For the reference sphere with a fingerprint, the RMS error between the measured height-data and the ones obtained from a fitted spherical model shows a significant reduction when compared before and after applying the procedure. In all corneal topography measurements in which Placido disk rings images were obscured by artefacts, the procedure showed significant improvement in both accuracy, as determined by Zernike polynomials fitting and the bootstrap mean-square error, as well as in the corneal coverage. The use of bootstrap-based techniques for assessing the applicability of the Zernike polynomial expansion to corneal topography data also showed that the proposed procedure leads to corneal estimates that are more reliable.

Although the procedure was evaluated in Medmont E300, it is by no means restricted to that particular instrument and can be used in all types of Placido disk videokeratoscopes. We chose Medmont E300 as a platform for our investigation because it has the option of exporting real topography data that is free of interpolation. Some problems

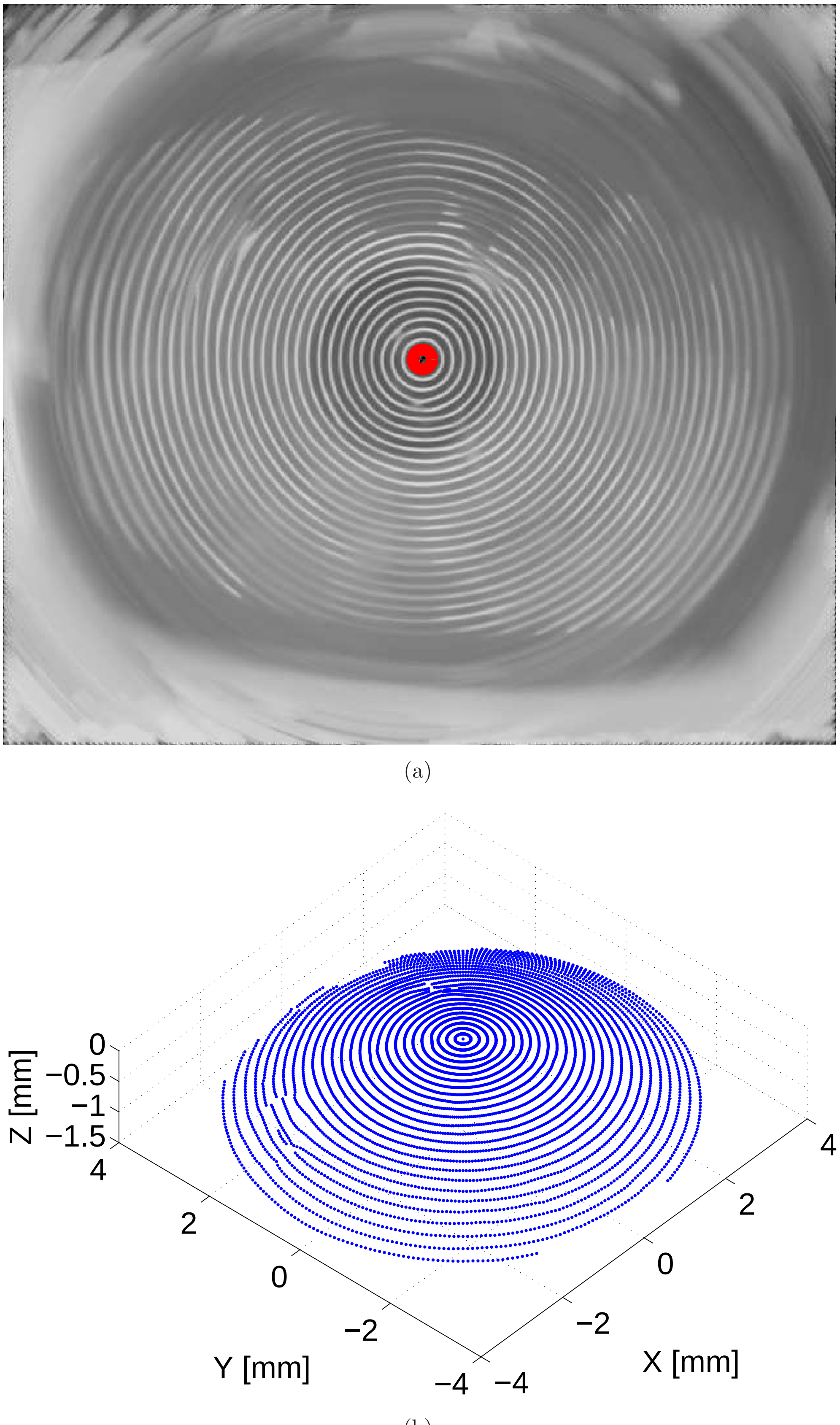


(a)



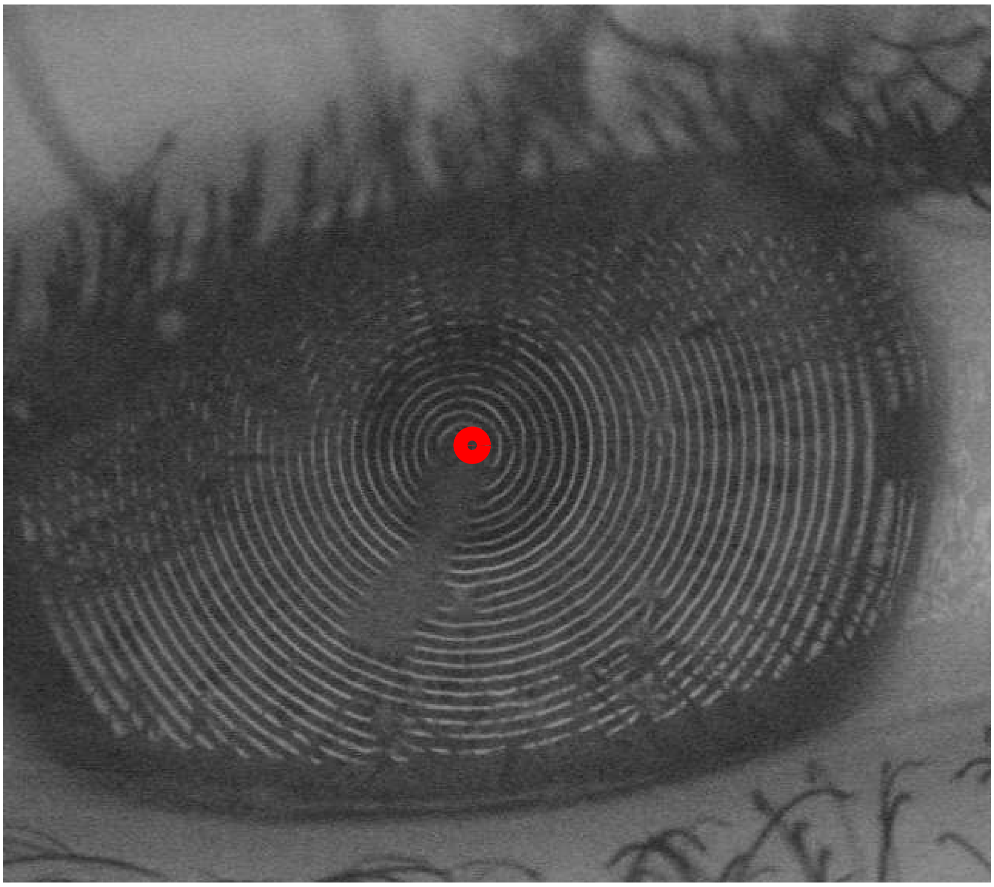
(b)

Figure 3.19. A non-standard acquisition case C (significant reflections from eyelashes, tear-film break-ups, and mucus): (a) the raw VKS image, and (b) its corneal topography.

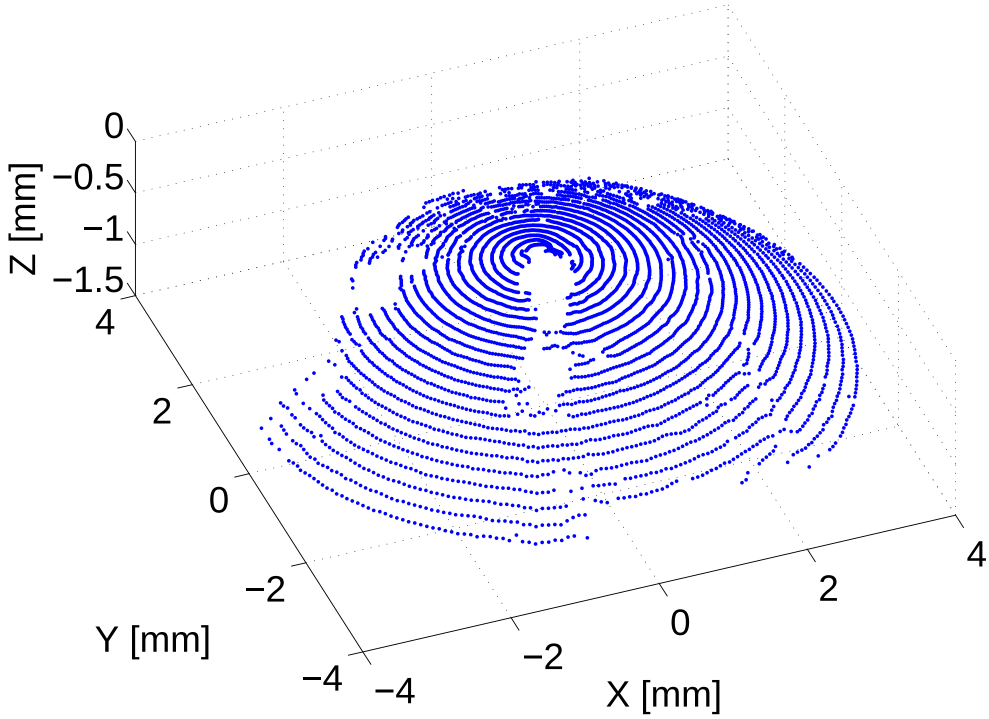

(a)

(b)

Figure 3.20. A non-standard acquisition case C (significant reflections from eyelashes, tear-film break-ups, and mucus): (a) the processed VKS image, and (b) its corneal topography.

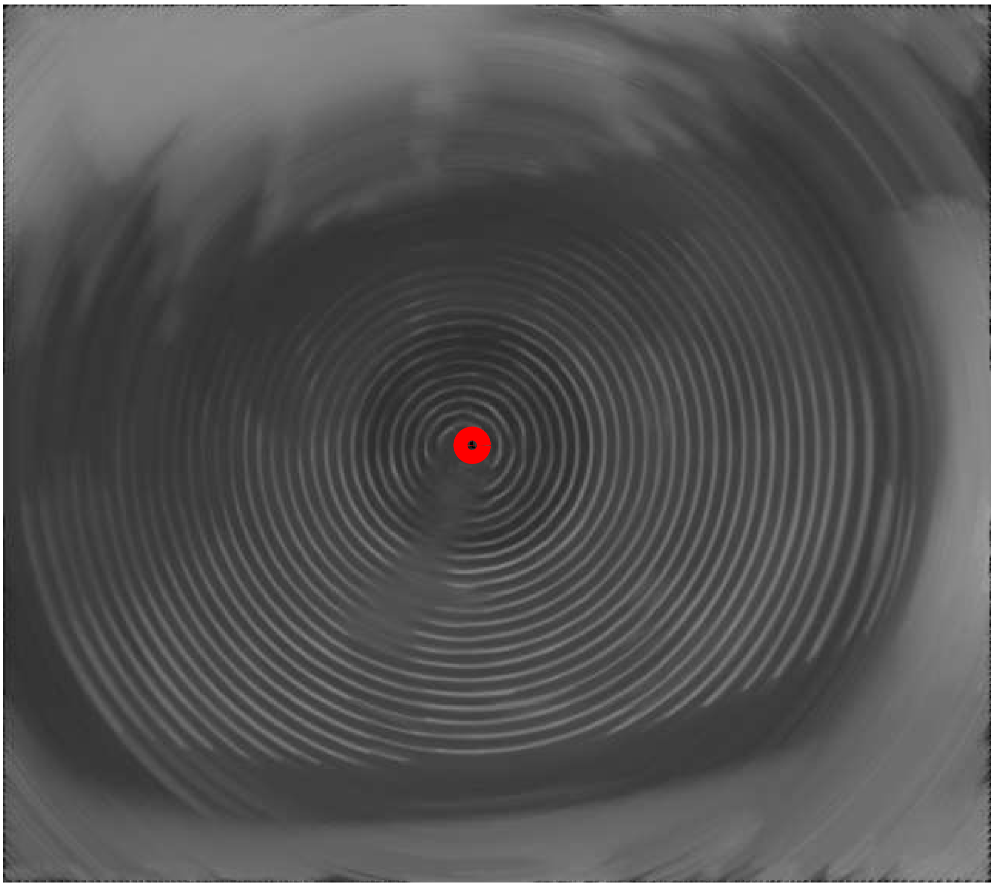


(a)

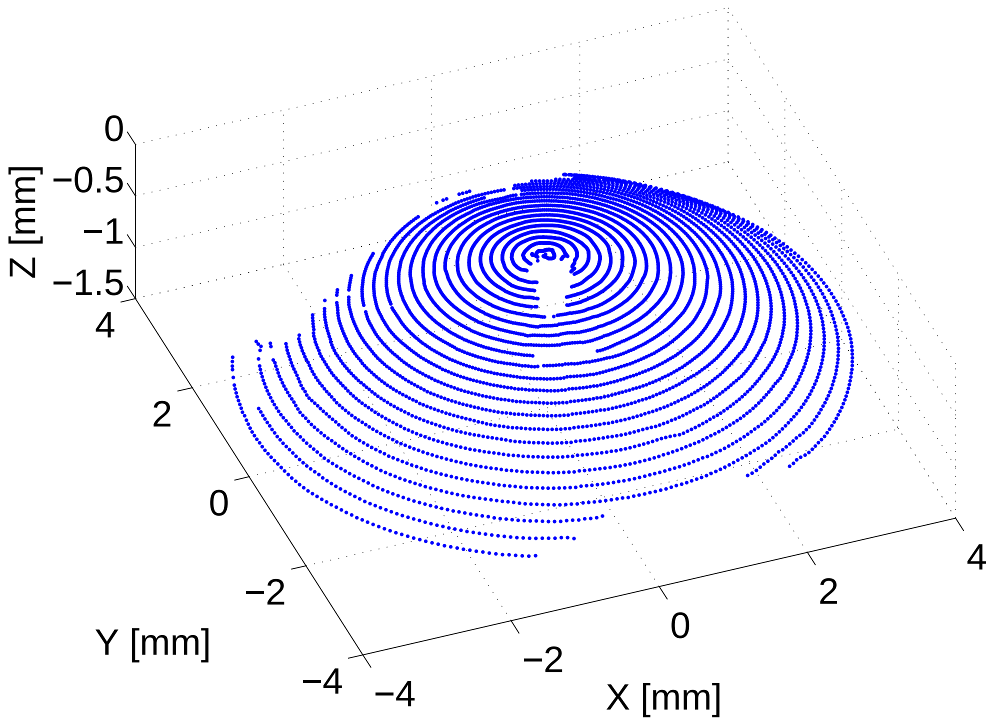


(b)

Figure 3.21. A non-standard acquisition case D (a large amount of mucus covers the central part of the cornea): (a) the raw VKS image, and (b) its corneal topography.



(a)



(b)

Figure 3.22. A non-standard acquisition case D (a large amount of mucus covers the central part of the cornea): (a) the processed VKS image, and (b) its corneal topography.

of missing observation can be rectified, to some extent, with interpolation techniques. They, however, are not able to recover large missing areas or significantly extend the corneal coverage as the proposed technique does.

Chapter 4

Model-Order Selection for Corneal Surfaces

4.1 Introduction and Motivation

The first step of image formation in the human eye occurs when the incident light is focused by the cornea. Therefore, clear vision highly depends on the optical quality of the corneal surface, which contributes to about 70% of the total optical power of the eye [26]. Normal, healthy cornea is a prolate surface that is steeper in the center and flatter in the periphery. Subtle changes in the corneal shape can result in a dramatic effect on the vision process [99].

Corneal shape reconstruction is typically measured with videokeratoscopes. In the most popular setting, a series of illuminated concentric rings, called the Placido disk pattern, are projected on the corneal surface and the rays reflected off the cornea are captured by a charge coupled device (CCD) video camera. Differential geometry methods (e.g., the arc-step algorithm [51]) are used to derive corneal topography by comparing the predefined geometry of the Placido disk pattern with the one obtained from the measured corneal surface.

Corneal modeling can be used as a tool for screening corneal diseases. Keratoconus, for example, distorts the corneal shape and results in a significant vision loss. Keratoconic patients should be screened for refractive surgeries because such treatments may worsen the corneal shape and lead to corneal transplantation. Hence, corneal modeling plays an essential role in diagnosing and managing keratoconus to assess subject's suitability for the treatment and prevent improper refractive surgeries [100]. Also, proper design and fit of contact lenses depend on the corneal topography.

Zernike polynomial expansions are commonly used for analyzing corneal wavefront error derived through ray tracing techniques [76]. This analysis is used to evaluate the corneal surface contribution to the eye's optics via the coefficients of the expansions that have optical interpretations (e.g., defocus and astigmatism). Moreover, Zernike polynomials expansion can be applied as a fitting tool to model corneal height-data, for example, for the sake of viewing small variations of the corneal surface [4]. There is, however, a continuous debate on the accuracy of modeling abnormal corneal surfaces using Zernike polynomials [5, 9]. In any case, one important issue that is encountered with the use of Zernike polynomials is the proper number of terms to be used (i.e., a model-order selection problem) and it is essential not to underestimate or overestimate the model-order.

Our work here has focused on two general approaches for model-order selection. First, an information theoretic-based approach and, second, a resampling-based approach. Information theoretic-based approaches include, but are not limited to, the Akaike's information criterion (AIC) and its corrected version ($\mathrm{AIC_C}$) [101,102], minimum description length (MDL) [103], Hannan–Quinn information criterion (HQC) [104], normalized maximum likelihood (NML) [105], conditional model-order estimation (CME) [106], Mallows' $\mathrm{C_P}$ criterion [107], and the efficient detection criterion (EDC) [108, 109]. Resampling-based approaches include the bootstrap [10, 13] and the "hook-and-loop" (HL) resampling plane [20, 21].

In [6], a bootstrap-based model-order selection procedure was proposed to estimate the number of Zernike terms for fitting a range of corneal surfaces. It was shown that the bootstrap procedure outperforms some information theoretic criteria, assuming independent and identically distributed (i.i.d.) realizations of measurement noise across the whole corneal surface. In a subsequent clinical study [7], the appropriateness of the bootstrap procedure for modeling normal corneas was shown. However, it was observed that the procedure underestimates the order of abnormal (e.g., keratoconic) corneal surfaces. In [8], this shortcoming was linked to the fact that in Placido disk videokeratoscopes corneal height-data does not fulfil the assumption of i.i.d. measurement noise [53], and a refined bootstrap procedure was proposed that took into account the spatial distribution changes in the measurement noise. Despite its superiority, the refined bootstrap-based method has not gained much popularity amongst vision researchers mainly due to its large computational complexity. Instead of using such a computationally expensive method, there is a need for a simple, but efficient, method that would deliver similar performance to the one based on the refined bootstrap.

The efficient detection criterion (EDC) was originally proposed in [108] for detecting the number of signals in white noise and its strong consistency was also proven. In [109], the EDC was used as a discriminant criterion for model and model-order selection in regression problems. While the EDC has the same general form of information theoretic criteria for model-order selection (i.e., a summation of a mean-square error function and a penalty function), it has the advantage of using non-fixed (i.e., flexible) penalty functions. This enables a wide range selection of penalty functions provided that they satisfy certain consistency conditions. Accordingly, the performance of an EDC based model-order selector can be improved by proper choice of the penalty function for the considered problem. Consequently, the EDC, which is to the best of our knowledge has not been used before for model-order estimation of corneal surfaces, is a a good candidate to substitute the method of refined bootstrap since its consistency, as we will show, can be controlled via one with-manageable-range parameter only.

In this Chapter, we compare the performance of the EDC against seven other information theoretic and two resampling-based criteria. We show, by simulation, the efficacy of using the EDC and illustrate its superiority to the other criteria in selecting the true model-order. Moreover, we assess its performance for real surfaces from different corneal groups and elucidate its ability to estimate model-orders that are in close agreement with clinical expectations and, yet importantly, to discriminate between different corneal groups.

This Chapter is organized as follows: Section 4.2 presents the methodologies for model-order selection for the Zernike polynomial expansion of a corneal surface. Section 4.3 shows the performance of different criteria on three simulated surfaces. Experimental results with performance analysis based on a range of real corneal surfaces are provided in Section 4.4, before a summary in Section 4.5.

4.2 Model-Order Selection

The corneal height-data in cylindrical coordinates (ρ_n, θ_n, S_n), $n = 1, 2, \ldots, N$, are typically modeled using a finite series of orthogonal Zernike polynomials[1]. This can be expressed, according to the single-indexed notation [73], as follows

$$S\left(\rho, \theta\right) = \sum_{k=1}^{K} a_k Z_k(\rho, \theta) + \varepsilon(\rho, \theta), \tag{4.1}$$

where $Z_k(\rho, \theta)$ is the single indexed k-th Zernike polynomial with a coefficient a_k and $\varepsilon(\rho, \theta)$ represents the measurement and modeling error. Eq. (4.1) can be written in a linear regression form

$$\boldsymbol{S} = \boldsymbol{Z}_k \boldsymbol{a}_k + \boldsymbol{\varepsilon}, \tag{4.2}$$

where $\boldsymbol{S}$ is an $N \times 1$ vector, $\boldsymbol{Z}_k$ is an $N \times k$ matrix of discrete, orthogonal Zernike polynomials, $\boldsymbol{a}_k$ is an $k \times 1$ vector of Zernike coefficients, $k = 1, 2, \ldots, K$, where K is the maximum number of Zernike polynomials available in Eq. (4.1), and $\boldsymbol{\varepsilon}$ is an $N \times 1$ vector of measurement and modeling errors. With such a given model,

---

[1] Details of using Zernike polynomials in modeling of corneal surfaces are found in Chapter 2.

the important task is to estimate the number, k, of Zernike polynomials required to model the corneal surface. This is a model-order selection problem, which is solved by computing a criterion score (an information theoretic or resampling-based), F, over a set of nested orders $\{k = 1, 2, \ldots, K\}$ and selecting the order that yields the minimum score according to

$$\hat{k}_o = \arg \min_{k \in \{1,2,\ldots,K\}} \{F(k)\}. \tag{4.3}$$

4.2.1 Information Theoretic Criteria

Information theoretic-based model-order selection criteria have a general form, in which the criterion score, F, is computed as a summation of two functions, G_1 (mean-square-error (MSE) function) and G_2 (penalty function) according to

$$F(k) = G_1\left(\hat{\sigma}_k^2, N\right) + G_2(k, N), \quad k = 1, 2, \ldots K, \tag{4.4}$$

where $\hat{\sigma}_k^2$ is the maximum-likelihood estimator of the variance of the surface residuals when k Zernike terms are fitted and is defined according to

$$\hat{\sigma}_k^2 = \frac{1}{N}\left(\boldsymbol{S} - \boldsymbol{Z}_k \hat{\boldsymbol{a}}_k\right)^T \left(\boldsymbol{S} - \boldsymbol{Z}_k \hat{\boldsymbol{a}}_k\right). \tag{4.5}$$

Here $(\cdot)^T$ denotes the transpose operator and

$$\hat{\boldsymbol{a}}_k = ({\boldsymbol{Z}_k}^T \boldsymbol{Z}_k)^{-1} {\boldsymbol{Z}_k}^T \boldsymbol{S}, \tag{4.6}$$

is the least-squares estimator [71] of the parameter vector $\boldsymbol{a}$ conditioned on k.

Many classical criteria have exactly the same MSE function, which is

$$G_1 = N \log(\hat{\sigma}_k^2). \tag{4.7}$$

Other criteria have slightly different ones, which are

$$G_1 = (N - k) \log(\hat{\sigma}_k^2), \tag{4.8}$$

$$G_1 = (N - k - 2) \log(\hat{\sigma}_k^2)/2, \tag{4.9}$$

or

$$G_1 = \hat{\sigma}_k^2 / \hat{\sigma}_K^2 - N. \tag{4.10}$$

Whilst these functions are slightly different, the difference between them, for the usual case in corneal surface modeling where the sample size, N, is relatively large, is small. Accordingly, we consider only the differences in the penalty functions as their behaviour have significant influence on the selected number of Zernike terms.

We compare the following criteria[2] for the penalty function G_2:

1. Akaike's Information Criterion (AIC) [101]:
$$G_2 = 2k. \tag{4.11}$$
2. Corrected Akaike's Information Criterion (AIC$_\mathrm{C}$) [102]:
$$G_2 = N\frac{1+k/N}{1-(k+2)/N}. \tag{4.12}$$
3. Minimum Description Length (MDL) [103]:
$$G_2 = k\log N. \tag{4.13}$$
4. Hannan–Quinn Criterion (HQC) [104]:
$$G_2 = 2ck\log\log N, \tag{4.14}$$
where $c \geq 1$.
5. Normalized Maximum Likelihood (NML) [105]:
$$G_2 = k\log(N\hat{R}) + (N-k-1)\log\left(\frac{N}{N-k}\right) - (k+1)\log k, \tag{4.15}$$
where $\hat{R} = \frac{1}{N}\hat{\boldsymbol{a}}_k^T\boldsymbol{Z}_k^T\boldsymbol{Z}_k\hat{\boldsymbol{a}}_k$.
6. Conditional Model-Order Estimation (CME) [106]:
$$G_2 = \frac{1}{2}\ln|\boldsymbol{Z}_k^T\boldsymbol{Z}_k| + \log\left(\frac{[\pi(N-k)]^{(N-k)/2}}{\Gamma\left(\frac{N-k}{2}\right)}\right),$$
where $\Gamma(\cdot)$ is Euler's Gamma function.
7. Mallows' C$_\mathrm{P}$ criterion [107]:
$$G_2 = \beta k, \tag{4.16}$$
where $\beta \geq 2$.

---

[2] More details about these criteria are found in Appendix A. However, for completeness and convenience, we repeat the part of the penalty functions here.

8. Efficient Detection Criterion (EDC) [108, 109]:

$$G_2 = kC_N, \tag{4.17}$$

where C_N satisfies

$$\lim_{N\to\infty} \left\{ \frac{C_N}{N} \right\} = 0, \tag{4.18}$$

and

$$\lim_{N\to\infty} \left\{ \frac{C_N}{\log\log N} \right\} = \infty, \tag{4.19}$$

to ensure strong consistency.

The HQC, the Mallows' $\mathrm{C_p}$, and the EDC can be considered as model-order selection criteria with non-fixed penalty functions. However, these functions have scalar parameters for the HQC and the Mallows' $\mathrm{C_p}$; c and β respectively, which can basically admit any value in the non-manageable range $[0, \infty]$. On the other hand, the penalty function of the EDC has a general realization; C_N, which can be controlled using one parameter in a manageable range.

For the EDC, there are many choices of C_N, which fulfil Eqs. (4.18) and (4.19). In [109], an *ad-hoc* method was used, which is in the form δN^γ, where $\gamma < 1$, but, in general, research related to selecting C_N is limited.

In this work, we investigate and compare the performance of two general realizations of C_N:

1. $C_{N_1} = N^\gamma$, $\gamma \in [0.1, 1]$. This agrees with [109] for $\delta = 1$.
2. $C_{N_2} = (N \log(N))^\gamma$, $\gamma \in [0.1, 1]$.

We select the realization C_{N_2} in order to maintain a faster divergence-to-infinity than C_{N_1} when both use the same γ values. This characteristic is useful as it allows the impact of different rates of divergence of the penalty function, on model-order selection for corneal surfaces, to be studied. It is worthy to note that this range of γ values yields explainable penalty functions. Selecting, for example, $\gamma = 0$ would provide a penalty function weaker than the one of the AIC (i.e., high probability of overestimation), while selecting $\gamma > 1$ would provide very strong penalty functions (i.e., high probability of underestimation).

4.2.2 Resampling-Based Criteria

The bootstrap-based model-selection criterion [10, 13] is used to approximate the bootstrap estimate of the prediction error[3]. In [6, 8], this criterion, and its refined version, have been used in the context of selecting the optimal number of Zernike terms for modeling corneal surfaces. In this case, Eq. (4.3) holds and

$$F(k) = \frac{1}{B} \sum_{b=1}^{B} SSE^{*}_{N,m}(k)_b, \tag{4.20}$$

where

$$SSE^{*}_{N,m}(k) = \frac{1}{N} \sum_{n=1}^{N} \Big(S(\rho_n, \theta_n) - \hat{S}^{*}(\rho_n, \theta_n) \Big)^2, \tag{4.21}$$

is the bootstrap sum of squared errors, B is the number of bootstrap resamples, and $\hat{S}^{*}(\rho_n, \theta_n)$ is an approximation to the bootstrap surface. The bootstrap surface is calculated from the residuals, which play an essential role in bootstrap-based model-order selection (for full details, see [6]).

We sample with replacement from

$$\sqrt{\frac{N}{m}} \left(\hat{\boldsymbol{r}} - \overline{\hat{\boldsymbol{r}}} \right), \tag{4.22}$$

where

$$\hat{\boldsymbol{r}} = \boldsymbol{S} - \boldsymbol{Z}_K \hat{\boldsymbol{a}}_K, \tag{4.23}$$

is the residuals vector computed after fitting the surface to the highest possible number of Zernike terms, and $\overline{\hat{\boldsymbol{r}}}$ is the sample mean of $\hat{\boldsymbol{r}}$. The parameter m is a scaling factor to ensure enough variability amongst different bootstrap surfaces and should satisfy [110]

$$\lim_{N \to \infty} m/N = 0, \tag{4.24}$$

and

$$\lim_{N \to \infty} m = \infty. \tag{4.25}$$

Finding the proper m value is a challenge and this problem becomes more difficult when the refined boostrap technique [8] is used, as m has to be set for each independent and identically distributed (i.i.d) partition of the residuals.

Recently, the "hook-and-loop" (HL) resampling plane [20, 21] was proposed to avoid the scaling problem. The core of the HL-based model selection is identical to the

---

[3] More details about resampling-based criteria are found in Appendix A. However, for completeness and convenience, we mention the relevant details here.

bootstrap-based one. The only difference is the replacement of the scaling parameter, m, with a chain of operations: detrending the residuals, sorting (ranking) them in ascending order, HL resampling,

$$\hat{r}^*_{(i)} = \frac{1}{2}\left(\hat{r}_{(i)} + \hat{r}_{(i+1)}\right) + \varphi_i, \tag{4.26}$$

where

$$\varphi_i \sim \mathcal{N}\left(0, \left[\frac{1}{6}\left(\hat{r}_{(i+1)} - \hat{r}_{(i)}\right)\right]^2\right), \tag{4.27}$$

and then re-sorting the residuals according to the indices of an ascending-ordered corneal height-data (in $\boldsymbol{S}$) (i.e., match small (large) residual values to small (large) height-data). The number of Zernike terms is then estimated using Eqs. (4.20) and (4.21).

4.3 Simulation Results

We tested the two considered general realizations of C_N in the EDC against other nine criteria, seven being information theoretic criteria and two being based on resampling (bootstrap and hook-and-loop). For this purpose, we created three arbitrary surface models defined on a unit circle. They are as follows:

$$S_1(\rho,\theta) = \frac{1}{2} + \frac{1}{2}Z_4(\rho,\theta) + Z_5(\rho,\theta) + \varepsilon(\rho,\theta), \tag{4.28}$$

which models a surface with the defocus term $(Z_4(\rho,\theta) = Z_2^0(\rho,\theta))$ and the astigmatism-with-axis-at-±45° term $\left(Z_5(\rho,\theta) = Z_2^{-2}(\rho,\theta)\right)$,

$$S_2(\rho,\theta) = \frac{1}{2}Z_7(\rho,\theta) + Z_8(\rho,\theta) + \varepsilon(\rho,\theta), \tag{4.29}$$

which models a surface with the coma-along-y-axis term $\left(Z_7(\rho,\theta) = Z_3^{-1}(\rho,\theta)\right)$ and the coma-along-x-axis term $(Z_8(\rho,\theta) = Z_3^1(\rho,\theta))$, and

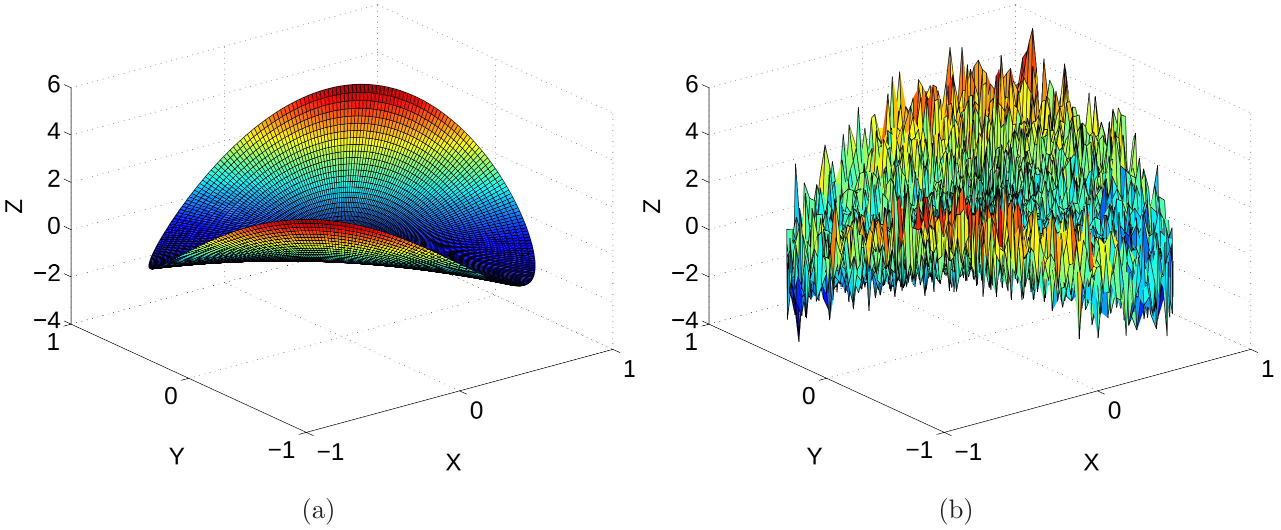
$$S_3(\rho,\theta) = Z_7(\rho,\theta) + Z_8(\rho,\theta) + Z_9(\rho,\theta) + Z_{10}(\rho,\theta) + \varepsilon(\rho,\theta), \tag{4.30}$$

which is $S_2(\rho,\theta)$ plus two trefoil terms, $\left(Z_9(\rho,\theta) = Z_3^{-3}(\rho,\theta)\right)$ and $(Z_{10}(\rho,\theta) = Z_3^3(\rho,\theta))$. This is a more-complex (i.e., 10-terms model) version of $S_2(\rho,\theta)$. Here $\rho \in [0,1]$ and $\theta \in [0,2\pi]$.

The measurement noise, $\varepsilon(\rho,\theta)$, describes the errors in estimating corneal height-data inherent to videokeratoscopy technique. This source of errors is different from the fitting residuals, $\hat{\boldsymbol{r}} = \boldsymbol{S} - \boldsymbol{Z}_k\hat{\boldsymbol{a}}_k$, as the latter occur regardless of the model used to describe the former. The measurement noise is modeled here as a zero-mean Gaussian process with variance increases with the values of the radial samples and is expressed as

$$\sigma^2 = \sigma_C^2 \left(\frac{\left[\rho_{(1)}, \rho_{(2)}, \ldots, \rho_{(n_\rho)}\right]}{\rho_{(n_\rho)}} \right)^2, \tag{4.31}$$

where σ_C^2 is a constant, $\left[\rho_{(1)}, \rho_{(2)}, \ldots, \rho_{(n_\rho)}\right]$ are the order statistics of the radial samples, $\rho_{(n_\rho)}$ is the maximum radial value, and n_ρ is the number of radial samples. In words, the variance of the measurement noise is zero at $\rho = 0$, increases in the radial direction, and gets its maximum value (i.e., σ_C^2) at the periphery of the surface. This spatial distribution of measurement noise is typical in Placido disk videokeratoscopes, as can be inferred from [53]. Figs. 4.1–4.3 depict surfaces $S_1(\rho,\theta)$, $S_2(\rho,\theta)$ and $S_3(\rho,\theta)$ along with their noisy versions using realizations of the measurement noise according to Eq. (4.31) with $\sigma_C^2 = 1$.



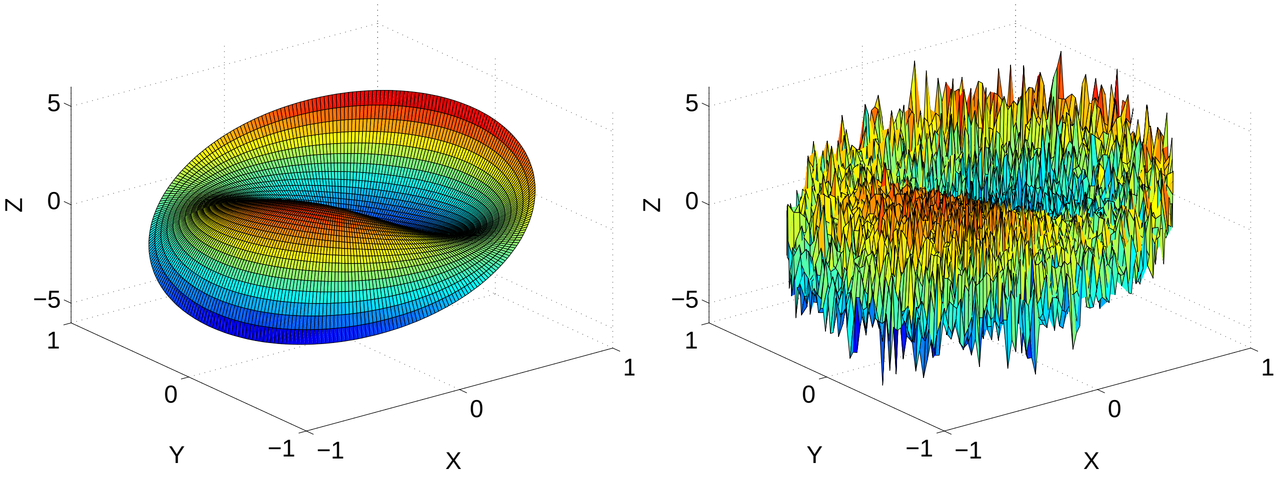
(a)

(b)

Figure 4.1. Surface $S_1(\rho,\theta)$ in (a) and its noisy version with $\sigma_C^2 = 1$ in (b).

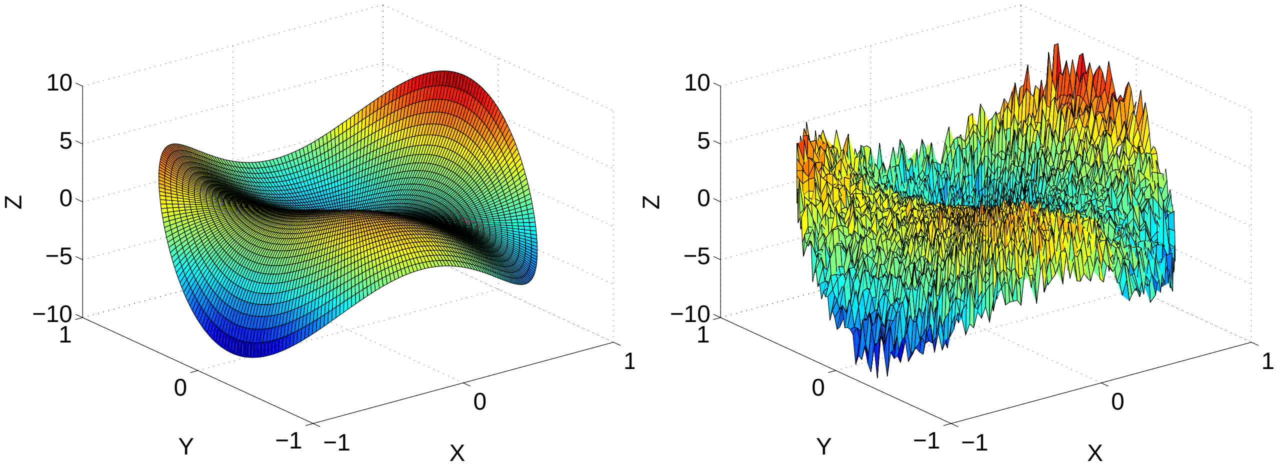
To select proper values of the parameter $\gamma \in [0.1, 1]$, for the realizations C_{N_1} and C_{N_2} of the EDC, and of the parameters c for the HQC and β for the Mallows' $\mathrm{C_p}$, we ran simulations on the model surfaces of Eqs. (4.28)–(4.30).

In Figs. 4.4–4.7, we show the empirical probabilities of selecting the correct model-order for each surface model ($k_{o_{S_1}} = 5$, $k_{o_{S_2}} = 8$, and $k_{o_{S_3}} = 10$) for EDC: kN^γ and EDC: $k(N\log(N))^\gamma$ using $\gamma = [0.1, 0.2, 0.3, \ldots, 1]$, HQC using $c = [2, 3, 4, \ldots, 15]$, and Mallows' $\mathrm{C_p}$ using $\beta = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 25, 50, 100]$, respectively. The following settings were used in the simulations: 1) four samples in the radial direction,



(a) (b)

Figure 4.2. Surface $S_2(\rho,\theta)$ in (a) and its noisy version with $\sigma_C^2 = 1$ in (b).



(a) (b)

Figure 4.3. Surface $S_3(\rho,\theta)$ in (a) and its noisy version with $\sigma_C^2 = 1$ in (b).

$\rho = [0.25, 0.50, 0.75, 1]$, and 36 equally spaced samples in the angular direction, $\theta = [0, (1)2\pi/35, (2)2\pi/35, \ldots, 2\pi]$, which correspond to a sample size of $N = 144$, (this is a small sample size for corneal modeling applications and was chosen to challenge the different model-order selection criteria), 2) $\varepsilon(\rho,\theta)$ was assumed according to Eq. (4.31) with $\sigma_C^2 = 1$, 3) the highest possible model-order was $K = 15$ (for $S_1(\rho,\theta)$ and $S_2(\rho,\theta)$) and $K = 20$ (for $S_3(\rho,\theta)$), and 4) $10,000$ independent Monte Carlo runs.

From Figs. 4.4 and 4.5, we observe that using γ values in the *middle* of the range $[0.1, 1]$ yields the highest probabilities of selecting the correct model-order. This outcome agrees with the fact that low γ values lead to weak penalty functions and ac-

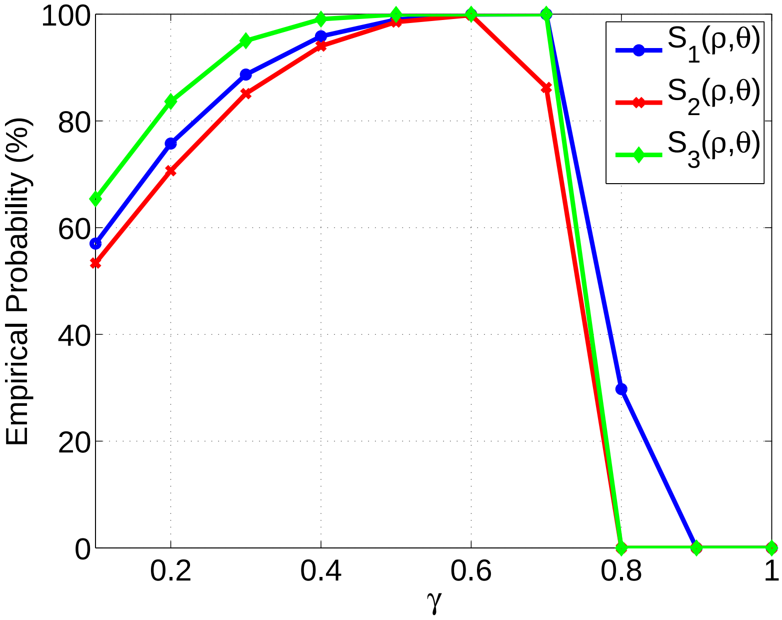


Figure 4.4. The empirical probabilities (in percent) of selecting the correct model-order for each surface model ($k_{o_{S_1}} = 5$, $k_{o_{S_2}} = 8$ and $k_{o_{S_3}} = 10$), evaluated over $10,000$ independent Monte Carlo runs for EDC: kN^{γ}, using $\gamma = [0.1, 0.2, 0.3, \ldots, 1]$, and assuming $\varepsilon(\rho, \theta)$ according to Eq. (4.31) with $\sigma_C^2 = 1$.

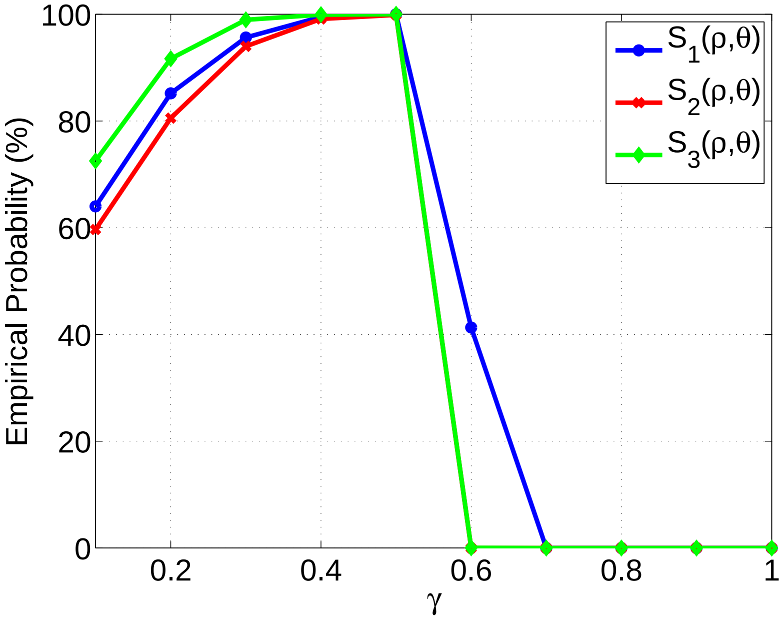


Figure 4.5. The empirical probabilities (in percent) of selecting the correct model-order for each surface model ($k_{o_{S_1}} = 5$, $k_{o_{S_2}} = 8$ and $k_{o_{S_3}} = 10$), evaluated over $10,000$ independent Monte Carlo runs for EDC: $(N \log(N))^{\gamma}$, using $\gamma = [0.1, 0.2, 0.3, \ldots, 1]$, and assuming $\varepsilon(\rho, \theta)$ according to Eq. (4.31) with $\sigma_C^2 = 1$.

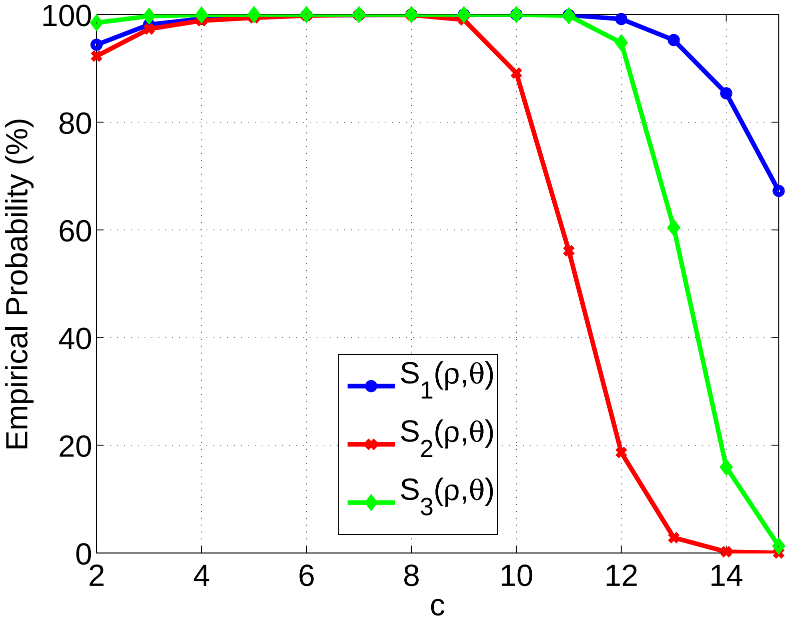


Figure 4.6. The empirical probabilities (in percent) of selecting the correct model-order for each surface model ($k_{o_{S_1}} = 5$, $k_{o_{S_2}} = 8$ and $k_{o_{S_3}} = 10$), evaluated over $10,000$ independent Monte Carlo runs for HQC using $c = [2, 3, 4, \ldots, 15]$, and assuming $\varepsilon(\rho, \theta)$ according to Eq. (4.31) with $\sigma_C^2 = 1$.

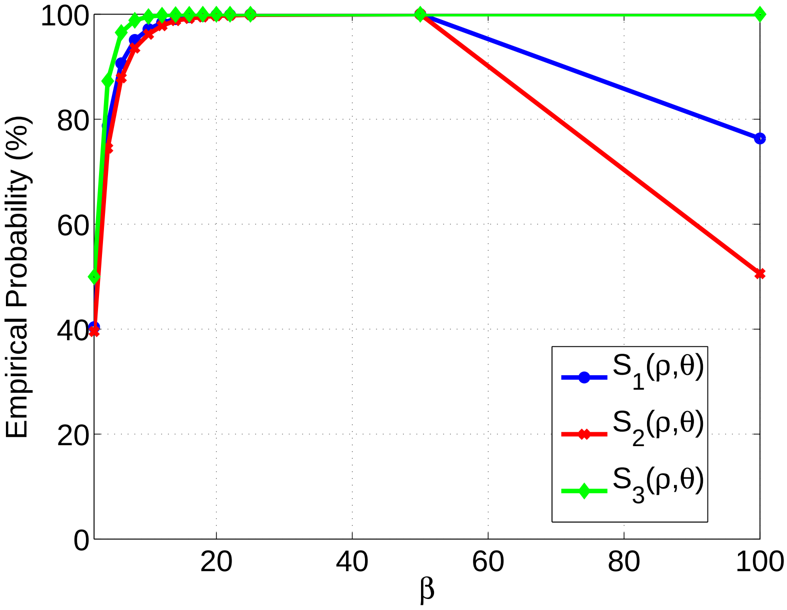


Figure 4.7. The empirical probabilities (in percent) of selecting the correct model-order for each surface model ($k_{o_{S_1}} = 5$, $k_{o_{S_2}} = 8$ and $k_{o_{S_3}} = 10$), evaluated over $10,000$ independent Monte Carlo runs for Mallows' C_p using $\beta = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 25, 50, 100]$, and assuming $\varepsilon(\rho, \theta)$ according to Eq. (4.31) with $\sigma_C^2 = 1$.

cordingly overestimating the model-order, and the opposite is true for high γ values. Therefore, the proper γ value can be selected to hold a compromise between the two extremes, underestimation and overestimation. On the other hand, we observe from Figs. 4.6 and 4.7 that the proper values of c and β, that yield high empirical probabilities, span large ranges. To select optimal values for γ, c, and β to proceed with, we allowed for each surface model a probability of error of 1% in $10,000$ Monte Carlo runs. Then, we selected the values of the parameters (i.e., γ, c, and β) that maximized the overall empirical probability of correct model-order estimation. The results were $\gamma = 0.6$ for EDC: kN^{γ}, $\gamma = 0.5$ for EDC: $k(N\log(N))^{\gamma}$, $c = 8$ for HQC, and $\beta = 50$ for Mallows' $\mathrm{C_p}$.

In Tables 4.1–4.3, we compare the empirical probabilities of underestimating, overestimating, and selecting the correct model-order for each surface model for the different information theoretic and resampling-based criteria described above. Besides the settings mentioned above, the following settings were used for the other criteria: 1) $B = 200$ for the bootstrap and the hook-and-loop procedures, and 2) $m = 12$ for the bootstrap procedure [6].

Table 4.1. Empirical probabilities (in percent) of underestimating, Pr[U], overestimating, Pr[O], and selecting the correct model-order, Pr[C], evaluated over $10,000$ independent Monte Carlo runs for $S_1(\rho,\theta)$ ($k_{o_{S_1}} = 5$) assuming $\varepsilon(\rho,\theta)$ according to Eq. (4.31) with σ_C^2 having two values; 0.25 and 1. The highest probabilities of selecting the correct model-order are indicated in bold fonts.

	$\sigma_C^2 = 0.25$			$\sigma_C^2 = 1$		
Criterion	Pr[U]	Pr[O]	Pr[C]	Pr[U]	Pr[O]	Pr[C]
MDL	0.0	8.8	91.2	0.0	9.0	91.0
AIC	0.0	34.6	65.4	0.0	33.6	66.4
$\mathrm{AIC_C}$	0.0	29.9	70.1	0.0	29.7	70.3
HQC: $c = 8$	0.0	0.0	**100.0**	0.0	0.0	**100.0**
NML	0.0	6.1	93.9	0.0	10.8	89.2
CME	0.0	9.6	90.4	0.0	18.1	81.9
EDC: $k(N\log(N))^{0.5}$	0.0	0.0	**100.0**	0.0	0.0	**100.0**
EDC: $kN^{0.6}$	0.0	0.1	99.9	0.0	0.1	99.9
BOOT	0.0	3.8	96.2	0.0	3.8	96.2
HL	73.5	3.3	23.2	100.0	0.0	0.0
$\mathrm{C_p}$: $\beta = 50$	0.0	0.0	**100.0**	0.0	0.0	**100.0**

It is clear from Tables 4.1–4.3 that the EDC criteria (based on penalty terms $k(N\log(N))^{0.5}$ and $kN^{0.6}$), HQC with $c = 8$, and Mallows' $\mathrm{C_p}$ with $\beta = 50$ give the best performance (almost 100 % in all cases). Note that the bootstrap-based procedure

Table 4.2. Empirical probabilities (in percent) of underestimating, Pr[U], overestimating, Pr[O], and selecting the correct model-order, Pr[C], evaluated over $10,000$ independent Monte Carlo runs for $S_2\,(\rho,\theta)$ ($k_{o_{S_2}} = 8$) assuming $\varepsilon(\rho,\theta)$ according to Eq. (4.31) with σ_C^2 having two values; 0.25 and 1. The highest probabilities of selecting the correct model-order are indicated in bold fonts.

	$\sigma_C^2 = 0.25$			$\sigma_C^2 = 1$		
Criterion	Pr[U]	Pr[O]	Pr[C]	Pr[U]	Pr[O]	Pr[C]
MDL	0.0	12.5	87.5	0.0	13.0	87.0
AIC	0.0	39.2	60.8	0.0	38.9	61.1
AIC_C	0.0	33.8	66.2	0.0	33.3	66.7
HQC: $c = 8$	0.0	0.0	**100.0**	0.0	0.1	**99.9**
NML	0.0	9.2	90.8	0.0	14.9	85.1
CME	0.0	13.1	86.9	0.0	21.9	78.1
EDC: $k(N\log(N))^{0.5}$	0.0	0.0	**100.0**	0.1	0.0	**99.9**
EDC: $kN^{0.6}$	0.0	0.2	99.8	0.0	0.2	99.8
BOOT	0.0	6.8	93.2	0.0	7.0	93.0
HL	17.4	10.9	71.7	94.5	0.5	5.0
C_p: $\beta = 50$	0.0	0.0	**100.0**	0.0	0.0	**100.0**

Table 4.3. Empirical probabilities (in percent) of underestimating, Pr[U], overestimating, Pr[O], and selecting the correct model-order, Pr[C], evaluated over $10,000$ independent Monte Carlo runs for $S_3\,(\rho,\theta)$ ($k_{o_{S_3}} = 10$) assuming $\varepsilon(\rho,\theta)$ according to Eq. (4.31) with σ_C^2 having two values; 0.25 and 1. The highest probabilities of selecting the correct model-order are indicated in bold fonts.

	$\sigma_C^2 = 0.25$			$\sigma_C^2 = 1$		
Criterion	Pr[U]	Pr[O]	Pr[C]	Pr[U]	Pr[O]	Pr[C]
MDL	0.0	3.6	96.4	0.0	4.1	95.9
AIC	0.0	27.3	72.7	0.0	26.8	73.2
AIC_C	0.0	19.9	80.1	0.0	20.2	79.8
HQC: $c = 8$	0.0	0.0	**100.0**	0.0	0.0	**100.0**
NML	0.0	1.1	98.9	0.0	3.0	97.0
CME	0.0	3.0	97.0	0.0	8.6	91.4
EDC: $k(N\log(N))^{0.5}$	0.0	0.0	**100.0**	0.0	0.0	**100.0**
EDC: $kN^{0.6}$	0.0	0.0	**100.0**	0.0	0.0	**100.0**
BOOT	0.0	4.1	95.9	0.0	4.2	95.8
HL	0.0	3.6	96.4	16.2	1.5	82.3
C_p: $\beta = 50$	0.0	0.0	**100.0**	0.0	0.0	**100.0**

yields less accurate but still reasonable results. On the other hand, the other information theoretic criteria overestimate the order due to the weakness of their penalty functions. For $S_3(\rho, \theta)$, more criteria perform well due to the used high SNR in this case, as the purpose of including this surface in the simulations was to examine the performance of the criteria that have non-fixed penalty functions for a higher-order surface model.

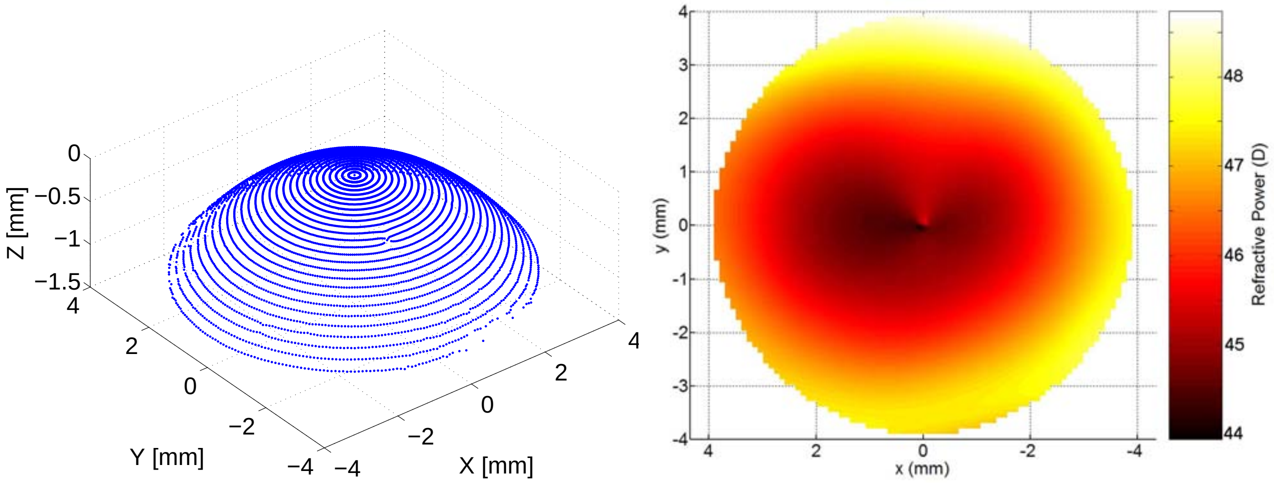
4.4 Experimental Results

To evaluate the potential of the EDC for estimating the number of Zernike terms for real corneas, we consider three different clinical groups:

1. normal corneas of 10 subjects,
2. astigmatic corneas (with cylindrical power greater than 0.75 Dioptres) of 12 subjects, and
3. keratoconic corneas of 20 subjects.

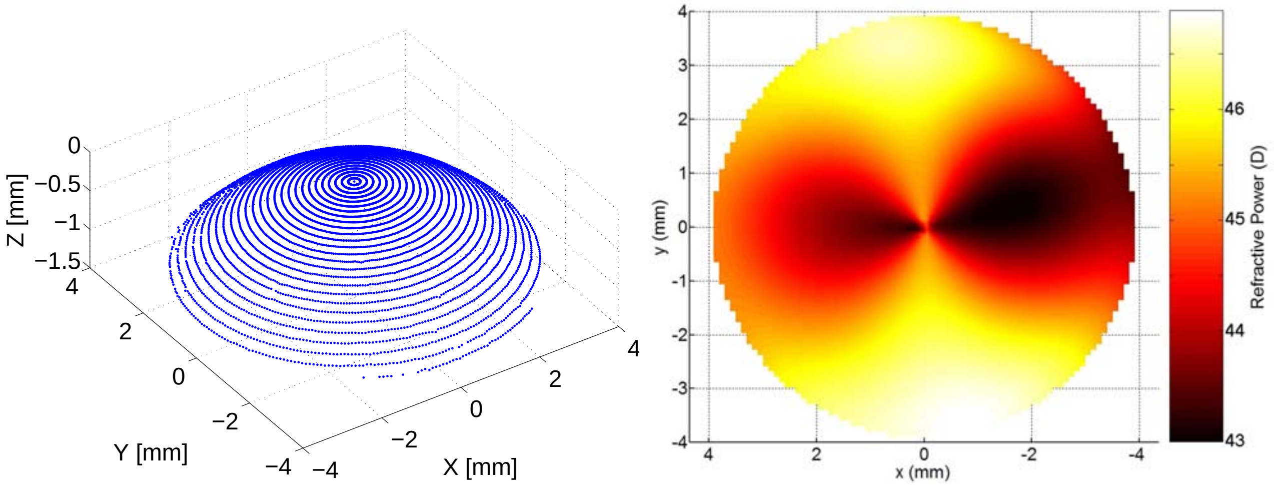
Only the right eyes of the subjects were considered. The keratoconic corneas had the average Steep K and Flat K values of (50 ± 5.2) and (45.5 ± 3.4) Dioptres, respectively, and surface regularity (SRI) and surface asymmetry (SAI) indices of (1.24 ± 0.42) and (5.12 ± 2.12), respectively. Corneal surface of at least 8 mm diameter was recorded for all of the subjects in the study. Figs. 4.8–4.10 show the corneal height-data and the refractive power representation of typical corneas for each of the considered clinical groups. The corneal height-data were acquired with Medmont E300 videokeratoscope (Medmont Pty Ltd, Melbourne, Australia). The Placido disk cone of the E300 videokeratoscope consists of 32 rings. The data was sampled in 300 equidistant semi-meridians.

We estimated the number of Zernike terms for 4-, 6- and 8 mm corneal diameters, yielding sample sizes in the range $4200 \leq N \leq 8916$. The highest possible model-order was set to $K = 120$, which corresponds to the 14th radial order. The scaling parameter for the bootstrap procedure was set to $m = 0.025$. This is a refined value of the one used in [6], of $m = 0.05$, to ensure the variability of the bootstrap residuals. This setting is related to an on-the-fly study of the noise power for the Medmont E300 videokeratoscope, which was found to be less than that of the Keratron videokeratoscope (used in [6] and [8]). The rest of the parameter settings remain as specified in Section 4.3.



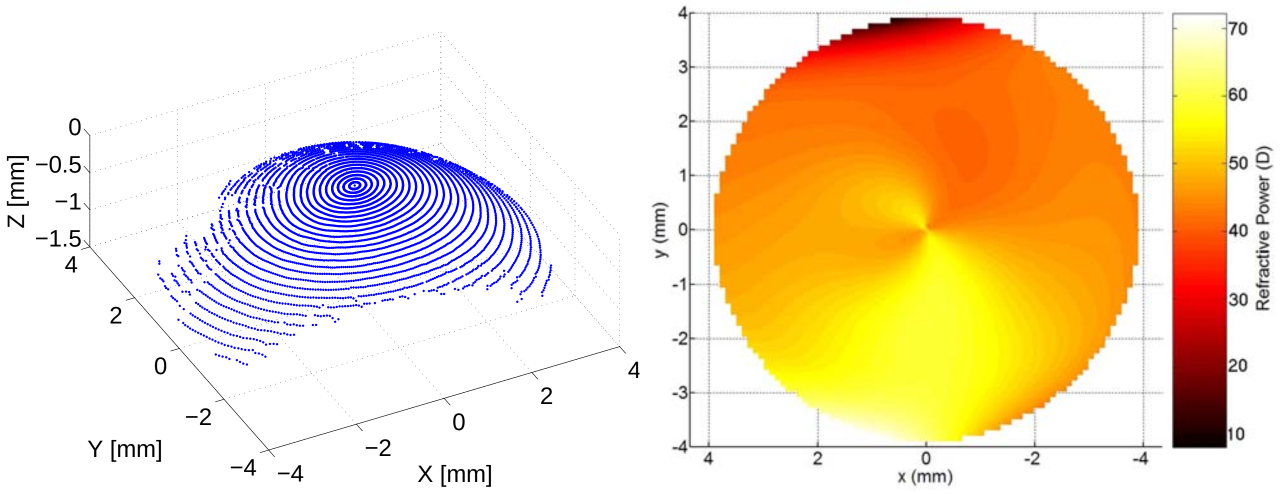
(a) (b)

Figure 4.8. Corneal height-data in (a) and corneal refractive power in (b) for an 8 mm diameter of a normal cornea.



(a) (b)

Figure 4.9. Corneal height-data in (a) and corneal refractive power in (b) for an 8 mm diameter of an astigmatic cornea.



(a) (b)

Figure 4.10. Corneal height-data in (a) and corneal refractive power in (b) for an 8 mm diameter of a keratoconic cornea.

In Table 4.4, the estimated number of Zernike terms from each model-order selection criterion for 4-, 6- and 8 mm diameters of the corneal surfaces considered in Figs. 4.8–4.10 are detailed. In general, the results indicate that the MDL, AIC, and AIC$_\mathrm{C}$ criteria overestimate the three considered surfaces since the estimated number of terms is close to the highest selected number of the Zernike terms (i.e., $K = 120$) across all corneal diameters. The limitations of these classical methods in the context of estimating the optimal number of Zernike terms for corneal surfaces have been previously noted [6]. These limitations result directly from the weakness of the penalty functions when compared to their corresponding MSE functions (both are summed up to the score function) in corneal surfaces modeling. On the other hand, the bootstrap and the HL clearly underestimate the astigmatic and the keratoconic surfaces. The rest of criteria have (almost) the trend of an increased estimated model-order with the complexity of the corneal surface. Figs. 4.11–4.13 illustrate the behavior of the score functions of the above criteria for 8 mm diameters of the corneal surfaces considered in Figs. 4.8–4.10. Note that EDC1 stands for EDC: $k(N\log(N))^{0.5}$, EDC2 stands for EDC: $kN^{0.6}$, besides that, for the overestimating criteria, MDL was only shown. Moreover, C$_\mathrm{P}$ was not shown because the dynamic range of its score is dramatically different than the scores of the other criteria, the logarithm score of bootstrap and HL were shown, and the score of the CME was scaled. Clearly, EDC-based procedures are able to find distinct minima while the other criteria perform poorly.

In the reminder of this Chapter, we only investigate and assess the performance of the NML, CME, EDC, HQC: $c = 8$, C$_\mathrm{P}$: $\beta = 50$, bootstrap and HL criteria. To have

Table 4.4. Estimated number of Zernike terms for the considered model-order selection criteria for 4-, 6- and 8 mm diameters of the corneal surfaces shown in Figs. 4.8–4.10.

	Diameter (mm)								
	Normal Cornea			Astigmatic Cornea			Keratoconic Cornea		
Criterion	4	6	8	4	6	8	4	6	8
MDL	120	119	104	120	117	105	120	120	120
AIC	120	119	120	120	120	120	120	120	120
$\mathrm{AIC_C}$	120	119	120	120	120	120	120	120	120
HQC: $c = 8$	90	32	55	52	52	45	119	117	119
NML	118	32	64	114	52	105	120	117	120
CME	105	32	55	52	52	55	120	117	119
EDC: $k(N \log(N))^{0.5}$	13	12	17	18	22	24	37	38	49
EDC: $kN^{0.6}$	14	12	17	18	22	24	47	54	49
BOOT	4	11	11	6	11	11	4	4	4
HL	11	12	12	9	12	11	4	6	6
$\mathrm{C_P}$: $\beta = 50$	90	32	54	52	52	45	116	113	119

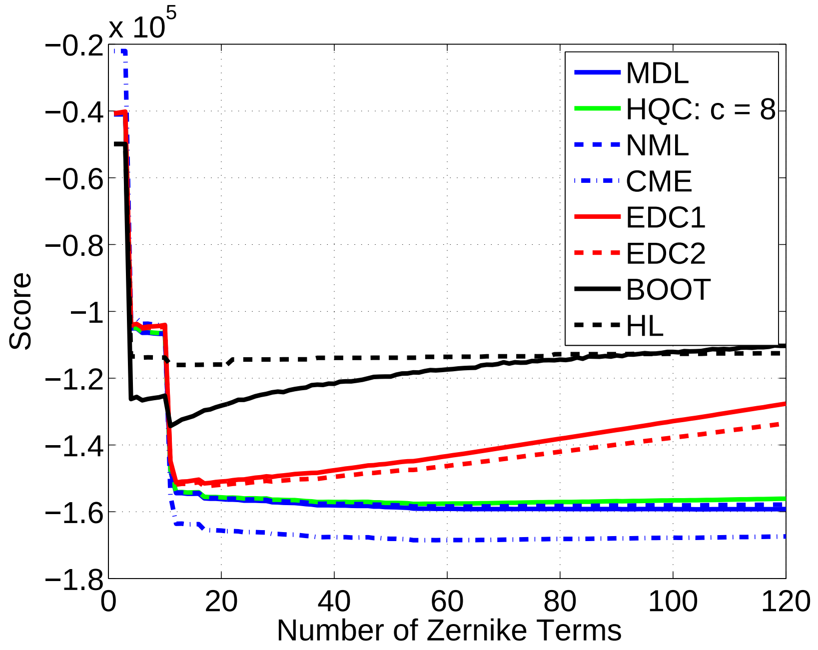


Figure 4.11. The behavior of the score functions of the considered information theoretic and resampling-based criteria for 8 mm diameters of the normal cornea considered in Fig. 4.8.

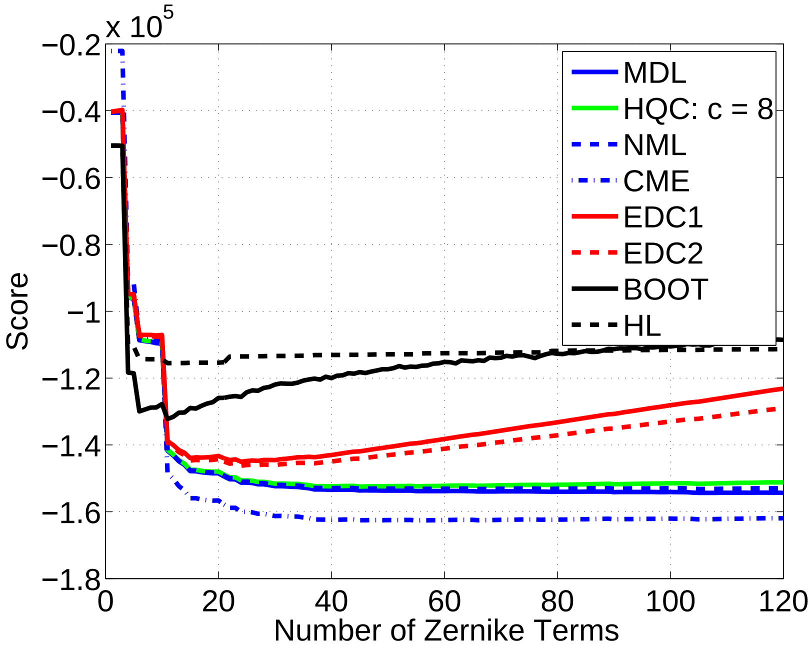


Figure 4.12. The behavior of the score functions of the considered information theoretic and resampling-based criteria for 8 mm diameters of the astigmatic cornea considered in Fig. 4.9.

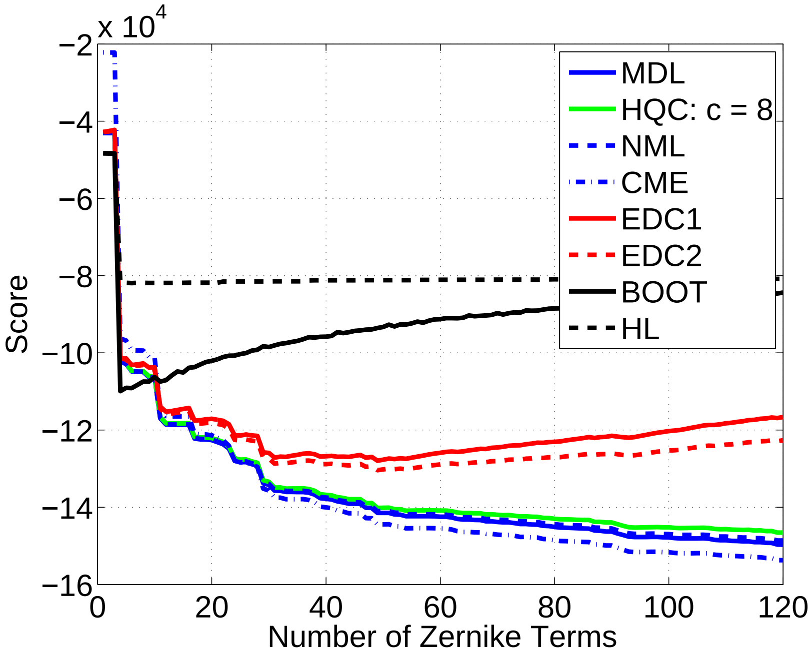
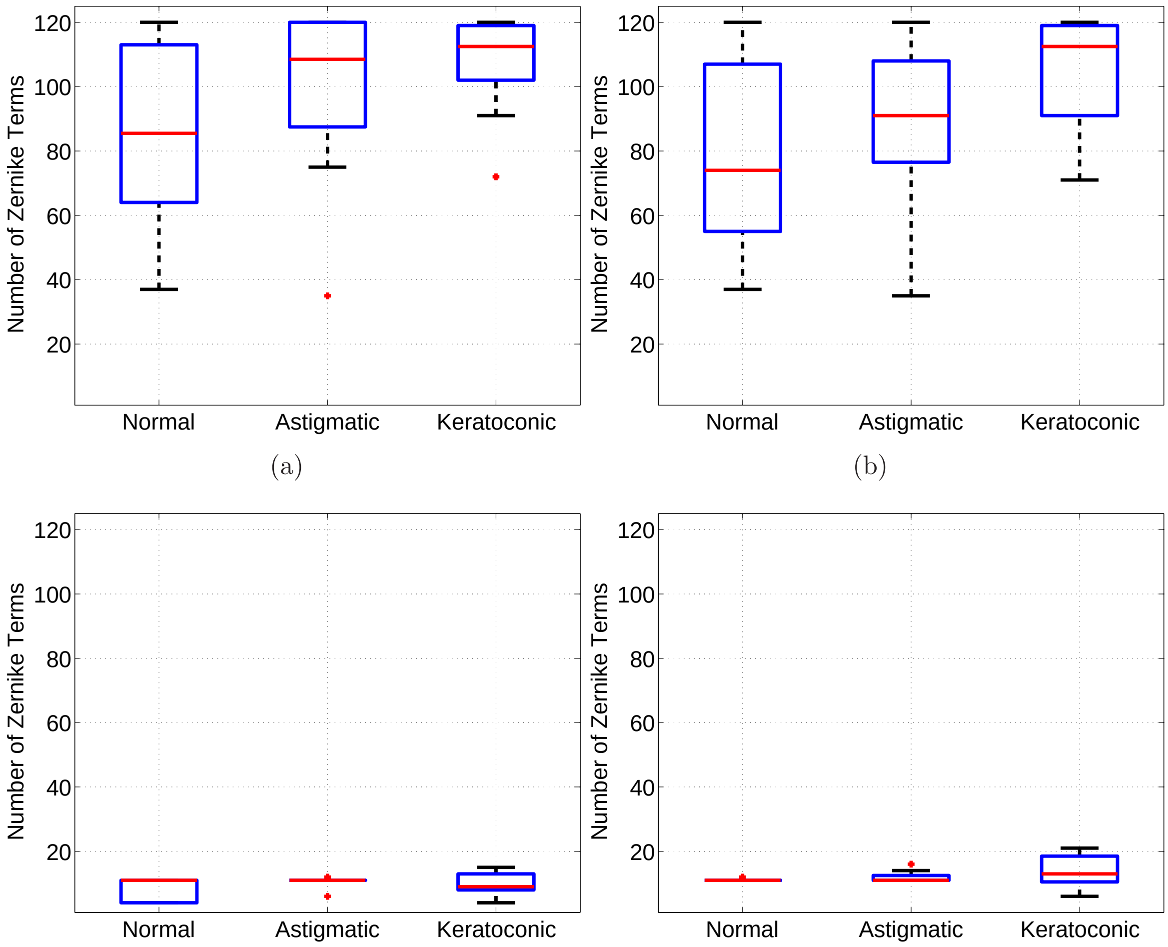


Figure 4.13. The behavior of the score functions of the considered information theoretic and resampling-based criteria for 8 mm diameters of the keratoconic cornea considered in Fig. 4.10.



(a) (b)

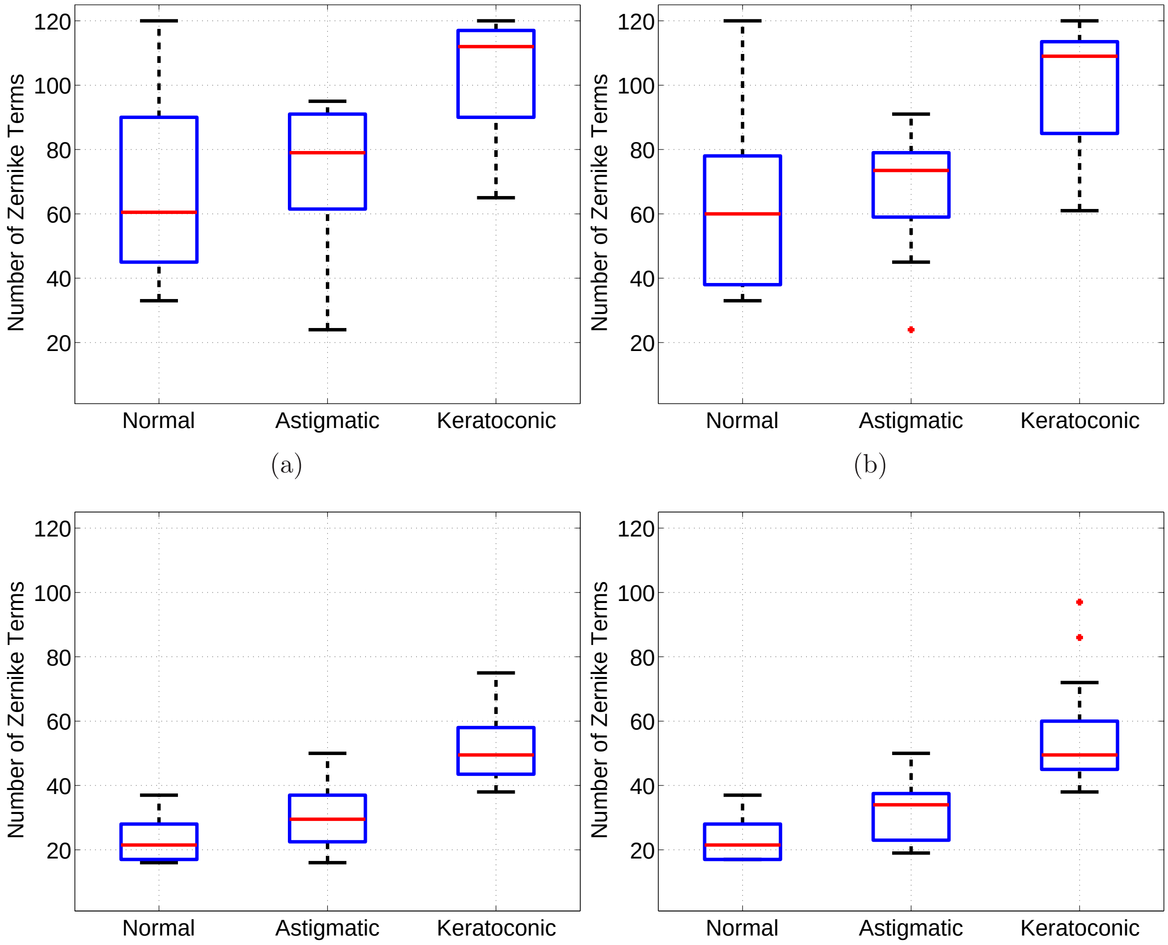
(c) (d)

Figure 4.14. Estimated number of Zernike terms for the three corneal groups (of 8 mm corneal diameter) for criteria: NML (a), CME (b), bootstrap (c), and HL (d).

a closer look to the performance of these criteria for different subjects in the three considered corneal groups, the boxplots of the estimated number of Zernike terms for 8 mm corneal diameters are depicted in Figs. 4.14 and 4.15 for the criteria with fixed penalty functions (i.e., NML, CME, bootstrap and HL), and for the criteria with non-fixed penalty functions (i.e., HQC, C_p, and EDC), respectively. Moreover, the sample mean of the estimated orders for 4-, 6- and 8 mm corneal diameters are shown in Fig. 4.16.

4.4.1 Corneal Discrimination

The number of subjects (observations) in each corneal group is limited (≤ 20). However, even if this number is large, the asymptotic distribution of the estimated model-



(a) (b) (c) (d)

Figure 4.15. Estimated number of Zernike terms for the three corneal groups (of 8 mm corneal diameter) for criteria: HQC: $c = 8$ (a), C_p: $\beta = 50$ (b), EDC: $k(N\log(N))^{0.5}$ (c), and EDC: $kN^{0.6}$ (d).

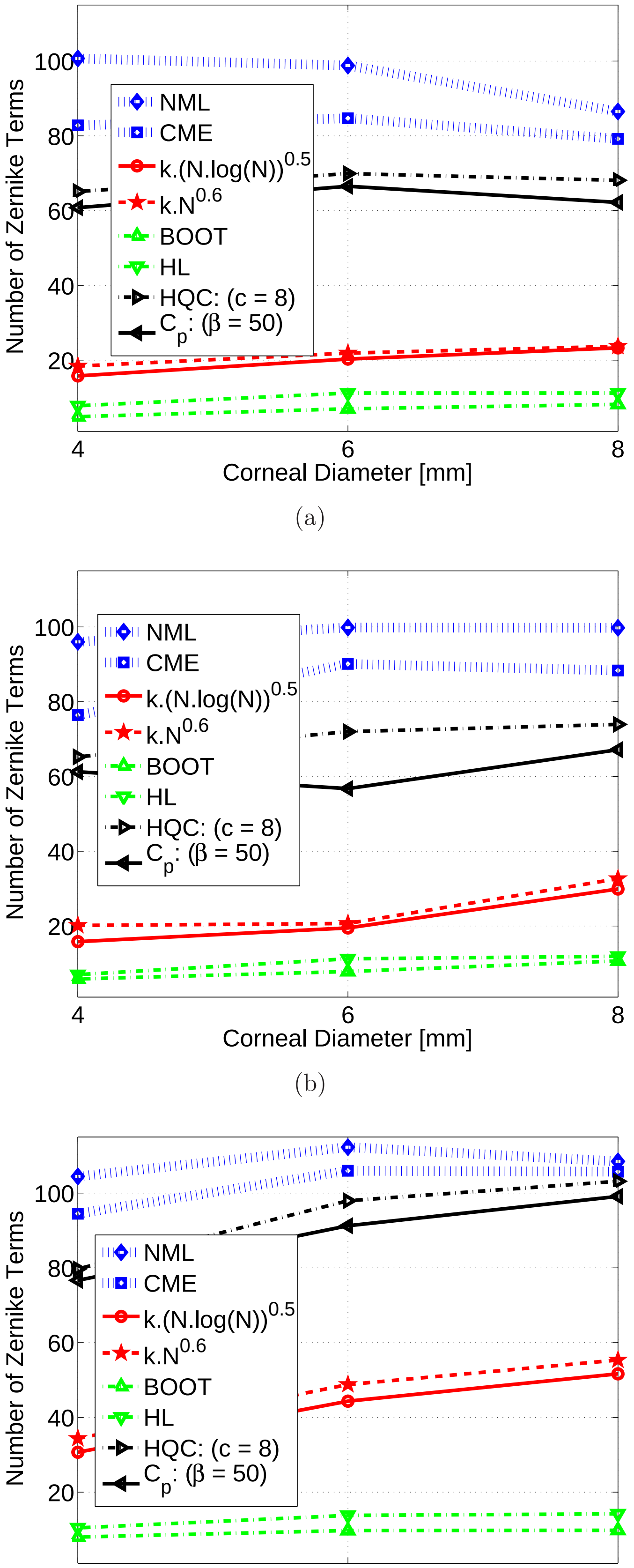
orders is unknown. In fact, experimental results, such as those from Figs. 4.14 and 4.15, suggest that the distribution of the estimated model-orders is not Gaussian. This motivates using nonparametric statistical hypothesis tests for evaluating the ability of the different criteria to discriminate between different corneal groups.

The Kruskal-Wallis (KW) test [111] is a nonparametric procedure for the one-way analysis of variance by ranks. We test the null hypothesis of the equality of the medians of the estimated model-orders (from each considered criterion) for the different corneal groups

$$H_0: \ \theta_1 = \theta_2 = \theta_3,$$

against the alternative

$$H_1: \ \text{Not } H_0.$$



(a)

(b)

(c)

Figure 4.16. Sample mean of the estimated orders at 4-, 6- and 8 mm corneal diameters for the normal (a), astigmatic (b), and keratoconic (c), corneal groups.

In the computation of the KW statistic, each observation is replaced by its tied rank in all observations. The KW statistic is defined as

$$T_{KW} = \frac{12}{D(D+1)} \sum_{j=1}^{J} \left[\frac{\left(\sum R_j\right)^2}{d_j} \right] - 3(D+1), \tag{4.32}$$

where $D = d_1 + d_2 + d_3 = 42$ is the total number of observations in the three corneal groups, $J = 3$ is the number of corneal groups, and $\sum R_j$ is the sum of the tied ranks in corneal group j. T_{KW} is usually divided by

$$A = 1 - \frac{\sum_{i=1}^{T} (t_i^3 - t_i)}{D^3 - D}, \tag{4.33}$$

where T is the number of sets of ties and t_i is the number of tied scores in the ith set of ties, as a tie correction.

If H_0 is true, T_{KW} is distributed approximately as a chi-squared distribution with two degrees of freedom. Thus, the KW test fails to accept the null hypothesis if

$$T_{KW} > \chi^2(\alpha; 2), \tag{4.34}$$

where α is the nominal significance level. We use the p-value[4] approach to measure the achieved significance level of the test

$$p\text{-value} = \Pr\left(T_{KW} \geq t_{KW} | H_0\right), \tag{4.35}$$

where t_{KW} is the realization of T_{KW} under H_0. The smaller the p-value, the stronger the evidence against H_0, which means that the considered model-orders represent different corneal groups.

In Table 4.5, the p-values computed by the KW test for the considered criteria, at 4-, 6-, and 8 mm corneal diameters are tabulated. The results indicate that EDC: $k(N\log(N))^{0.5}$ and EDC: $kN^{0.6}$ yield the smallest p-values for all considered corneal diameters. However, the p-values for the other criteria are small too. Carrying out multiple comparison test reveals that EDC: $k(N\log(N))^{0.5}$ and EDC: $kN^{0.6}$ are the criteria, which result in order medians of the normal and the astigmatic groups being significantly different from the keratoconic group for all considered corneal diameters.

---

[4] The p-value is the smallest probability that yields a rejection of the null hypothesis. Once the p-value has been calculated, one can decide a significance level, α, to either accept or reject the null hypothesis.

Table 4.5. p-values computed by the KW test for the considered criteria at 4-, 6- and 8 mm corneal diameters. The smallest p-values achieved for each column are indicated in bold fonts.

Criterion	Corneal Diameter [mm] 4	6	8
NML	0.21	0.05	0.11
CME	0.06	0.01	0.03
EDC: $k(N\log(N))^{0.5}$	**0.00**	**0.00**	**0.00**
EDC: $kN^{0.6}$	**0.00**	**0.00**	**0.00**
BOOT	**0.00**	0.06	0.25
HL	0.01	0.66	0.18
HQC: $c = 8$	0.28	**0.00**	**0.00**
Cp: $\beta = 50$	0.14	**0.00**	**0.00**

4.4.2 Population Match

We use the nonparametric Wilcoxon signed-ranks test [112] in order to determine whether or not the estimated model-orders of each corneal group (for each considered criterion) is derived from a population, in which the median, θ, is equal to a specified value. In [8], it was observed that for 8 mm corneal diameters, the normal cornea required a fifth radial-order (21 Zernike terms) model while the distorted corneas required an eighth radial-order (45 Zernike terms) model. Taking into consideration that the bootstrap procedure usually underestimates the distorted corneal surfaces, we test the following hypotheses

Normal corneas:

$$\begin{aligned} H_{01}: &\quad \theta_1 = 21, \\ H_{11}: &\quad \theta_1 \neq 21. \end{aligned} \tag{4.36}$$

Astigmatic corneas:

$$\begin{aligned} H_{02}: &\quad \theta_2 = 28, \\ H_{12}: &\quad \theta_2 \neq 28. \end{aligned} \tag{4.37}$$

Keratoconic corneas:

$$\begin{aligned} H_{03}: &\quad \theta_3 = 55, \\ H_{13}: &\quad \theta_3 \neq 55. \end{aligned} \tag{4.38}$$

The procedure of the exact-table Wilcoxon test starts with computing the difference between the estimated model-orders and the assumed median value, $Y = \hat{k}_o - \theta$. The absolute value of Y is then sorted (ranked) according to

$$R = |Y_{(1)}| < |Y_{(2)}| < \ldots < |Y_{(d)}|, \tag{4.39}$$

after averaging of tied scores and with zero terms excluded. Here d is the number of observations in the considered corneal group. Two rank sums are computed:

$$\begin{aligned} W^+ &= \sum_{i=1}^{d} \eta_i^+ R, \\ W^- &= \sum_{i=1}^{d} \eta_i^- R, \end{aligned}$$

where $\eta_i^+ = I\left(Y_{(i)} > 0\right)$, $\eta_i^- = I\left(Y_{(i)} < 0\right)$, and $I(\cdot)$ is the indicator function. The Wilcoxon test statistic is then defined as

$$T_W = \min\{W^+, W^-\}, \tag{4.40}$$

and p-values are thus calculated for T_W from tables.

In Table 4.6, the p-values computed by the Wilcoxon test for the considered criteria at 8 mm corneal diameters are tabulated. To compare the performance of different γ values for the EDC, we include the p-values for other less consistent γ values, which are 0.4 and 0.6 for EDC: $k(N\log(N))^\gamma$, and 0.5 and 0.7 for EDC: kN^γ. It is observable that EDC: $k(N\log(N))^{0.5}$ and EDC: $kN^{0.6}$ provide the strongest evidence for not rejecting the null hypothesis for all corneal groups. Therefore, any of them can be used as a superior criterion for selecting the number of Zernike terms for each corneal group, noting that EDC: $k(N\log(N))^{0.5}$ provides, on average, a stronger evidence than EDC: $kN^{0.6}$. These results agree with previous research based on the bootstrap [8]. However, the EDC has the advantage over the bootstrap approach as there is no need to set the m value for each partition of the residuals and, importantly, the computational requirements are dramatically lower.

4.5 Summary

Finding the optimal number of Zernike polynomials to model the corneal surface using data from a videokeratoscope is non trivial and requires a systematic model-order selection procedure such as information theoretic or resampling-based criteria.

Table 4.6. p-values computed by the Wilcoxon test for the considered criteria at 8 mm corneal diameters. The largest p-values achieved for each column are indicated in bold fonts.

Criterion	Normal	Astigmatic	Keratoconic
NML	0.00	0.00	0.00
CME	0.00	0.00	0.00
EDC: $k(N\log(N))^{0.4}$	0.00	0.00	0.00
EDC: $k(N\log(N))^{0.5}$	**0.44**	**0.61**	0.09
EDC: $k(N\log(N))^{0.6}$	0.02	0.00	0.00
EDC: $kN^{0.5}$	0.00	0.00	0.00
EDC: $kN^{0.6}$	0.34	0.13	**0.44**
EDC: $kN^{0.7}$	0.24	0.00	0.00
BOOT	0.00	0.00	0.00
HL	0.00	0.00	0.00
HQC: $c = 8$	0.00	0.00	0.00
C_p: $\beta = 50$	0.00	0.00	0.00

The coefficients of a Zernike polynomial expansion to a corneal surface tend to dramatically decrease with order causing a problem for the majority of information theoretic criteria whose penalty functions are too weak to correctly estimate the model-order. On the other hand, resampling-based techniques either underestimate the model-order or are computationally very expensive. An alternative is to seek criteria in which the penalty function can be arbitrarily changed. The efficient detection criterion (EDC) was a good candidate for investigation as its penalty function is controllable by one parameter.

We have evaluated the performance of EDC-based model-order selection techniques and benchmarked them against seven information theoretic-based criteria and two resampling-based techniques. The superiority of the EDC based criteria was shown with simulated surfaces. For real corneal surfaces, these criteria result in models that are in closer agreement with clinical expectations. Also, it was demonstrated that they have good ability in distinguishing normal corneal surfaces from astigmatic and keratoconic surfaces.

It is noteworthy to mention that a parameter value around the middle of the considered range (i.e., 0.6 for kN^{γ} or 0.5 for $k(N\log(N))^{\gamma}$) is the most reasonable choice for the divergence rate of the penalty function when modeling corneal surfaces is investigated. This provides moderate amplitude of the penalty function versus the MSE function and, hence, enables the EDC criterion to find the minimum score. Finally, the EDC based model-order selection techniques are computationally efficient.

On the other hand, we have carried out simulations to select the proper parameters for other information theoretic criteria that have non-fixed penalty functions (e.g., β for the Mallows' $\mathrm{C_p}$ criterion and c for the HQC). Then, these values were used for model-order selection of corneal surfaces. However, we found that the EDC still outperforms these criteria due to the difficulty of selecting optimal values for their parameters in the corresponding range, $[0, \infty]$, versus the manageable parameter's range of the EDC, $[0.1, 1]$.

Having such a significant performance of the EDC, it is worthy to be investigated for modeling other sources of corneal data such as corneal slope and curvature, which are the initial source of data that can be retrieved by arc-step based videokeratoscopes. It is believable that these types of corneal data are less prone to errors than corneal height-data as there is no need for subsequent integration processes. Therefore, more accurate modeling can be expected. However, this modeling is behind the scope of this thesis.

Chapter 5
Conclusions and Future Work

Two interrelated problems, the enhancement of videokeratoscopic images and its subsequent accurate estimation of corneal topography and the estimation of the proper number of Zernike terms to be fitted to corneal height-data (i.e., model-order selection), have been addressed in this work.

First, the problem of accurate estimation of corneal topography (via corneal height-data) for non-standard image acquisition scenarios was addressed. Basically, three-dimensional corneal height-data is reconstructed using differential geometry algorithms from two-dimensional corneal images that are captured using corneal topographers. The state-of-the-art topographers (e.g., videokeratoscopes), that depend on Placido disk pattern, assume a uniform pre-corneal tear-film to enable the proper formation of the reflected pattern, and a wide ocular aperture in order to capture as much corneal area as possible. However, in some cases like subjects with dry-eye syndromes, Asian eyes, severe reflections from eyelashes, kids/children, and/or combination of such cases, these requirements are not fulfilled and consequently yield low-quality videokeratoscopic images. Accordingly, inaccurate corneal topography is reconstructed. One question of interest was the ability of achieving an accurate corneal topography estimation from such deteriorated videokeratoscopic images by integrating an image-enhancement procedure in the corneal reconstruction algorithm of a commercial videokeratoscope, the Medmont E300.

Second, optimal model-order selection of corneal surfaces when expanded using orthogonal Zernike polynomials, was addressed. Here, optimality means closer agreement with clinical expectations. The traditional approach for estimating the number of Zernike polynomials depends on the clinician's experience. Therefore, it was interesting to investigate whether this approach can be done automatically, using systematic model-order selection approaches, within an acceptable precision.

The two problems are interrelated in the sense that accurate and reliable corneal modeling, whatever the used modeling approach was (e.g., Zernike polynomials or any other type of functional modeling), highly depends on the reconstructed corneal surface that is essentially conditioned on the quality of the captured videokeratoscopic image.

In the following, a summary including conclusions of the achieved work for the two addressed problems is provided in Section 5.1 while an outlook to future work is given in Section 5.2.

5.1 Summary and Conclusions

5.1.1 Enhancement of Videokeratoscopic Images for Accurate Corneal Topography Estimation

Enhancing videokeratoscopic images that were captured in non-standard scenarios was treated in Chapter 3. The main objective was to develop an improved image processing technique that overcomes the errors that might happen in the image acquisition stage, due to missing or incomplete data, in order to improve the estimation of corneal topography.

Since the image processing techniques used in commercial videokeratoscopes are proprietary, there is a limited number of studies that addressed the enhancement issue. They basically interpolate the missing data from its neighborhood provided that the number of missing data points are limited. However, this procedure does not work properly for each case, particularly when there is a large area of occluded conreal surface (i.e., by mucus) or when large parts of the Placido disk patten are fused due to corneal irregularities or instability of the pre-corneal tear-film.

To come with an efficient algorithm for enhancing the videokeratoscopic image, some ground truth should be provided. We decided to use the videokeratoscope's own reconstruction algorithm for achieving this task. That is, proposing and applying an adaptive image processing algorithm for enhancing videokeratoscopic images and then assessing its accuracy by comparing the reconstructed corneal surfaces before and after applying the enhancement algorithm.

The proposed enhancement algorithm was essentially a combination of a linear filter, the adaptive pixel-wise Wiener filter, and a non-linear filter, depends on mathematical morphology. This filtering was achieved after transforming the videokeratoscopic image to the polar representation to ease the processing as the Placido disk pattern tends to be a quasi-line pattern in the polar domain. After processing has been done, the videokeratoscopic image is transformed back to the Cartesian domain. The superiority[1], and hence the suitability, of this combination for the application at hands was in terms of preserving the edges of the Placido disk pattern unaltered (using the Wiener filter stage that yields less smoothing at the edges' locations) and keeping the overall

---

[1] We have tried other types of smoothing filters rather than Wiener filter. Basically, simple approaches work but, on the other hand, the Wiener filter was superior due to using local statistical moments for processing of different image regions.

geometry of the pattern while only closing/bridging the missing data points (using the morphological closing operator).

To assess its accuracy, the proposed enhancement algorithm was used to process a wide range of videokeratoscopic images, for both artificial surfaces (e.g., a calibrated sphere smeared with a fingerprint) and real corneas from different subjects and different irregularity conditions. For all cases, the enhancement algorithm has provided images that were subjectively of much higher quality than the input images in terms of recovering missing data, correcting for the fused parts of the Placido disk pattern, and, therefore, increasing the corneal coverage area. However, to objectively assess the quality of the processed videokeratoscopic images, expanding the reconstructed corneal surfaces from the pre- and the post-processed images with the same number Zernike polynomials was done. Then, the root-mean-squared error and the relative increment in corneal surface coverage were calculated and used as accuracy measures. Reduction in the root-mean-squared error and increment in corneal surface coverage were registered for all considered videokeratoscopic images. Moreover, and yet importantly, it is worthy to assess the deterioration, if any, that the enhancement algorithm would cause for the original corneal topography. To measure this factor, estimating the proper number of Zernike terms was done for the pre- and post-processed corneal surface. Results indicated that the enhancement algorithm provides the same number of Zernike terms for both surfaces for most cases and in other cases, it gives more realistic number of Zernike terms that agree with clinical expectations. Therefore, the proposed algorithm yields more accurate corneal topography estimation and at the same time does change the original topography of the corneal surface.

Finally, it is noteworthy to mention that although the proposed enhancement procedure was evaluated for the Medmont E300 videokeratoscope, it is usable in all types of Placido disk videokeratoscopes [16].

5.1.2 Model and Model-Order Selection

Model-order selection for corneal surfaces when fitted using orthogonal Zernike polynomials was addressed in Chapter 4. Moreover, since intensive use of model and model-order selection criteria was performed during the course of this work, different criteria were studied in detail and tested using simple polynomials to highlight their corresponding behaviors, in Appendix A.

Generally, there are many different approaches that can be used for model and model-order selection. We focused in this work on two approaches only, information theoretic based criteria and resampling based criteria. Information based theoretical criteria include, but are not limited to, the Akaike's information criterion (AIC) and its corrected version (AIC_C), minimum description length (MDL), Hannan–Quinn information criterion (HQC), normalized maximum likelihood (NML), conditional model-order estimation (CME), Mallows' C_P criterion, and the efficient detection criterion (EDC). Resampling-based approaches include the bootstrap and the "hook-and-loop" (HL) resampling plane. There are many factors that affect the consistency of model selection like the sample size, signal-to-noise ratio, the parameters' estimation algorithm, and the rank of the design matrix. However, all model selection criteria can be seen as a compromise between overestimation (i.e., overfitting or variance) and underestimation (i.e., underfitting or bias). This compromise can be controlled in some criteria that have *non-fixed* penalty functions like HQC, Mallows' C_P, and EDC. Moreover, the consistency of the bootstrap can be controlled using the sub-sample length or the scaling parameter, m.

For all simulations in Appendix A, we have considered variety of sample sizes and different signal-to-noise ratios, while full-rank design matrices were used in all cases. Rank deficiency was ignored on purpose in order to concentrate only on the performance of different model selection criteria and, moreover, to cope with the nature of the design matrices (i.e., matrices of Zernike polynomials), used in the experimental results of Chapter 4, which were always full-rank.

Specifically, in Appendix A, comparing the performance of the biased- versus the unbiased parameters' estimation used in the model selection criteria was achieved. Biased estimation was performed by scaling the unbiased parameters vector in order to minimize its mean squared error. The results indicated that using biased parameters' estimation-based model selection outperforms the unbiased one, in some cases. Interestingly, using biased estimation has resulted in a significant increment in the consistency of the conditional model-order estimation criterion (CME) for one of the simulation examples [19]. On the other hand, we have investigated using partial-model selection search rather than nested search for model-order selection. it was shown via simulations that the performance of all considered model-order selection criteria improves significantly using such a search method. Moreover, it was shown that combining the bootstrap with the partial-model selection search provides results that significantly outperform other information theoretic based criteria, particularly at low signal-to-noise ratios [18].

For model-order selection of corneal surfaces, which was performed in Chapter 4, a performance comparison of the penalty functions of ten different selection criteria (both information-theoretic based and resampling-based) has been done. Simple surface models in noise and real corneal surfaces from different clinical groups were investigated in the comparison. It was shown via surface models that the efficient detection criterion (EDC) outperforms all other considered criteria. It is noteworthy to mention that a parameter value around the middle of the considered manageable range, $[0.1, 1]$, (i.e., $\gamma = 0.6$ for kN^{γ} or $\gamma = 0.5$ for $k(N\log(N))^{\gamma}$) is the most reasonable choice for the divergence rate of the penalty function. On the other hand, the parameters of other considered criteria that have *non-fixed* penalty functions (e.g., c for Hannan–Quinn information criterion (HQC) and β for Mallows' $\mathrm{C_p}$) can basically have any values in the range $[0, \infty]$. Therefore, there is a large uncertainty about the proper values of such parameters when used in the context of modeling corneal surfaces. To achieve a fair comparison, the parameters that yielded the highest consistency of EDC, HQC, and Mallows' $\mathrm{C_p}$ for the surface models were used to select the model-order of real corneas. Assessing the performace of these criteria was in terms of their ability to discriminate between different corneal groups (via analysis of variance test) and the closeness of the selected orders to clinical expectations (via signed-ranks test). Experimental results have shown that the EDC outperforms all other criteria in this context.

In summary, using the EDC: $k(N\log(N))^{0.5}$ for model-order selection of corneal surfaces provides moderate amplitude of the penalty function versus the mean squared error function and, hence, enables the EDC criterion to find the minimum score. This was not possible when using other information theoretic criteria either due to the uncertainty of the proper parameter values in criteria with *non-fixed* penalty functions or due to the weakness of the *fixed* penalty functions in other criteria. Finally, the EDC based model-order selection techniques are computationally efficient, which is the contrary of the resampling-based techniques [17].

5.2 Future Work

5.2.1 Enhancement of Videokeratoscopic Images for Accurate Corneal Topography Estimation

Following the work in [16], an adaptive scheme for selecting the size of the structure element of the morphological closing operator and/or the length of the Wiener filter can be used to replace the fixed ones. This scheme should depend on the local changes in the Placido rings' pattern. Keep in mind that we need a classification stage in order to separate the outliers that might occur due to tear-film breaks and/or mucus covers the corneal surface from those irregularities that originally belong to the corneal shape itself. The frame work of model-order selection can be used in this context as these proper sizes and/or filter lengths can be selected in such a way to keep a trade-off between the amount of introduced smoothness and a measure of the change in the original corneal topography. However, for doing so, ground-truth surfaces with different complexities and exactly known surface topographies are required.

Having the achieved improvement in corneal topography estimation in the static mode (i.e., from still-images) [16], the proposed enhancement algorithm can be extended and applied in the dynamic mode (i.e., using a sequence of videokeratoscopic images). This can be used to track the dynamics of the corneal topography (e.g., using Kalman filters) and, at the same time, to provide an accurate corneal topography estimation. Moreover, the dynamics in the precorneal tear-film can be more accurately estimated and separated from the irregularities in the original corneal topography (if any).

Rather than applying the proposed enhancement algorithm in the image/video domain, it can be applied directly on the corneal height-data. In this way, the proprietary image processing techniques, used in commercial topographers to extract the Placido disk pattern, can be relaxed and considered irrelevant to the enhancement procedure. Therefore, more accurate assessment of the enhancement algorithm is possible. Depending on the achieved accuracy, the enhancement algorithm can be combined with the arc-step algorithm.

5.2.2 Model and Model-Order Selection

Following the work of [18, 19], biased estimation and/or partial-model order selection search can be used for model-order selection of corneal surfaces. It is worthy to mention here that the computational complexities of such techniques are large when it comes to modeling corneal surfaces, in particular in the case of using partial-model order selection search. However, simulation results have shown the superiority of such a searching method and, therefore, it is worthy to be applied for corneal modeling especially in cases with low signal-to-noise ratios.

Rather than using least squares estimation, least absolute shrinkage and selector operator (LASSO) techniques can be used for model-order selection of corneal surfaces. Moreover, robust estimation techniques can be investigated for compensating for the outliers in the videokeratoscopic images.

On the other hand, different model-order selection criteria can be compared in the context of modeling the radial profile of corneal surfaces, which yields a direct estimation of the radial order. This can be seen as one-dimensional modeling with small sample size (e.g., 33 points) and, therefore, the techniques of biased estimation and partial-model order selection search [18, 19] can be applied and investigated with almost the same degree of computational complexity found in the simulation examples, where sample sizes of 30 points were used.

Having such a significant performance of the efficient detection criterion (EDC) for modeling corneal height-data [17], it is worthy to be investigated for modeling other sources of corneal data such as corneal slope and/or curvature, which are the initial sources of data that can be retrieved by arc-step based videokeratoscopes. It is believable that these types of corneal data are less prone to errors than corneal height-data as there is no need for subsequent integration processes. Therefore, more accurate modeling can be expected.

Appendix A: Model and Model-Order Selection

A.1 Introduction and Motivation

Model selection is the task of selecting a statistical model to describe a set of observations (i.e., data) from a set of competitive models. On the other hand, model-order selection is the task of estimating the dimension (i.e., number of parameters) of such statistical models. Generally, the model should be *complex* enough to explain relations in the data but, on the other hand, *simple* enough to be understood and used. In terms of hypothesis testing, model selection can be considered as a multiple composite hypothesis testing [113]. A number of issues should be kept in mind when considering model and model-order selection problems.

1. Models are approximations as the true model is always unknown.
2. Model selection is a bias-variance trade-off that is a balance between simplicity (i.e., underfitting or underestimation) and complexity (i.e., overfitting or overestimation). Simplicity implies fewer parameters to estimate and, therefore, less variability but larger bias. On the other hand, complexity means more parameters to estimate and, therefore, more variability but less bias.
3. Irrelevant parameters can be excluded from the model, which is the the principle of parsimony.

Model and model-order selection related problems arrive in many forms and on widely varying occasions. Applications include, but are not limited to, parametric spectrum estimation [114], system identification [115], radar [116], polynomial phase signal modeling [117], neural networks [118] and biomedical engineering [6, 16, 17, 82].

One way to solve the problem of model and model-order selection is to assume a prior probability density function (PDF) for the unknown parameters of each investigated model (i.e., the Bayesian approach [119]). Then, the maximum *a posteriori* (MAP) decision rule [120] can be used. However, in practice, prior PDFs are not always available. Accordingly, many model selection rules have been formulated based on other concepts to overcome the limitation of the Bayesian approach. We focus here on two general approaches. First, an information theoretic-based approach and, second, a resampling-based approach. Information theoretic-based criteria include, but are not limited to, the Akaike's information criterion (AIC), the corrected Akaike's information criterion ($\mathrm{AIC_C}$) [101, 102], minimum description length (MDL) [103], Hannan–Quinn information criterion (HQC) [104], normalized maximum likelihood (NML) [105], Mallows' $\mathrm{C_p}$ criterion [107], and the efficient detection criterion (EDC) [108,109]. These criteria have been generally derived based on asymptotic behaviour. Therefore, their performance is limited for finite data records. Conditional model-order estimation (CME) [106] was derived to account for finite data records by basing the decision on sufficient and ancillary statistics. Resampling-based approaches include the bootstrap [10, 13] and the "hook-and-loop" (HL) resampling plane [20, 21]. They are based on the residuals to approximate the prediction errors.

For linear regression modeling, which is our concern in this thesis, least square estimation (LSE) [71] is usually used to estimate the parameters conditioned on the chosen model. For the case of Gaussian noise, the LSE estimator coincides with the minimum variance unbiased (MVU) estimator [71]. On the other hand, biased estimators have recently shown superior performance as compared to MVU estimators in view of minimizing the MSE of an estimator [121]. In [19], biased estimation was applied in the context of model and model-order selection and its superiority was demonstrated in some cases.

Having the advantage of the sparsity property used for model selection, the approaches of model selection and model-order selection can be combined. This merging technique yields partial-model selection search methodology that can be used for improving the consistency[2] of model-order selection criteria. It was shown in [18] that partial-model order selection search significantly improves the performance of many model-order selection criteria, particularly at low signal-to-noise (SNR) values.

In this Appendix, the motivation is to highlight the differences between a wide range of model and model-order selection criteria in the context of modeling regression polynomials. Accordingly, more details about the criteria used in this thesis are provided. Moreover, aspects related to biased estimation and partial-model selection search and their performance improvements over traditional methods are included. Simulation results are given in order to clarify the different concepts.

---

[2] Consistency is related to the ability of the model and/or model-order selection method to identify the true (or, more precisely, the closest-to-true in some sense) model that generated the data.

A.2 The Model

Consider a linear model of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{w}, \tag{A.1}$$

where $\boldsymbol{y}$ is an $N \times 1$ observations vector, $\boldsymbol{X}$ is an $N \times k$ fixed full-rank design matrix, $\boldsymbol{\theta}$ is an $k \times 1$ unknown parameters vector, $k < N$, and $\boldsymbol{w}$ is an $N \times 1$ iid noise vector of Gaussian distribution with zero mean and $\sigma^2\boldsymbol{I}$ covariance matrix or $\boldsymbol{w} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$.

The model in Eq. (A.1) can be used in either modeling the whole corneal hight-data when fitted using Zernike polynomials[3] or for modeling linear regression polynomials, which is the concern in this Appendix.

In such a model of Eq. (A.1), two fundamental problems arise. First, the question to ask is how to estimate the number of the parameters, k, which is the model-order selection problem. Second, whether there is a subset of $\{1, 2, \ldots, k\}$ that will best (in some sense) represent the data, which is the model selection problem.

Let ψ, which we call henceforth the model, be a subset of $\{1, 2, \ldots, k\}$. Subsequently, let $\boldsymbol{X}_\psi$ be a sub-matrix of $\boldsymbol{X}$ and $\boldsymbol{\theta}_\psi$ be a sub-vector of $\boldsymbol{\theta}$, both containing the indices of the non-zero components in ψ. Then, a model corresponding to ψ is [12]

$$\boldsymbol{y} = \boldsymbol{X}_\psi\boldsymbol{\theta}_\psi + \boldsymbol{w}. \tag{A.2}$$

The aim of model selection is to choose, or more precisely estimate, model ψ so that it best represents the data. This can be done using a full- or a partial-search methodology [18, 19], as we will explain below.

The least squares estimation (LSE) [71] is used to estimate the model parameters

$$\hat{\boldsymbol{\theta}}_\psi = \left(\boldsymbol{X}_\psi^T\boldsymbol{X}_\psi\right)^{-1}\boldsymbol{X}_\psi^T\boldsymbol{y}, \tag{A.3}$$

where $(\cdot)^T$ denotes the transpose operator.

Different model selection criteria can be investigated for estimating the model and/or the model-order. However, there is a continuous difficulty in selecting an appropriate model selection technique in practice as some criteria work quite well in some examples while they completely fail in other examples.

---

[3] Using regression models for corneal surfaces was provided in Chapter 4.

A.3 Information-Based Model Selection Criteria

Information theoretic-based model and model-order selection criteria have a general form, which selects the model (or model-order) that minimizes a "score function", calculated as

$$F(\psi) = N \log \hat{\sigma}_{\psi}^{2} + G(k_{\psi}, N), \tag{A.4}$$

where $\hat{\sigma}_{\psi}^{2}$ is the maximum-likelihood estimator of the variance of the residuals (for model ψ)

$$\hat{\sigma}_{\psi}^{2} = \frac{1}{N}\left(\boldsymbol{y} - \boldsymbol{X}_{\psi}\hat{\boldsymbol{\theta}}_{\psi}\right)^{T}\left(\boldsymbol{y} - \boldsymbol{X}_{\psi}\hat{\boldsymbol{\theta}}_{\psi}\right), \tag{A.5}$$

and $\hat{\boldsymbol{\theta}}_{\psi}$ is calculated as shown in Eq. (A.3).

In Eq. (A.5), the term $G(k_{\psi}, N)$ is the non-negative penalty function, a function in the number of parameters in the model, k_{ψ}, and the number of data points, N. It increases with the number of parameters in the model. Moreover, it takes different forms according to different information-based criteria. On the other hand, the term $N \log \hat{\sigma}_{\psi}^{2}$ decreases as the number of parameters increases. With these two terms, model and model-order selection techniques hold the balance between overestimation and underestimation. However, this form of *fixed penalty term* yields a very good model-order selection performance in one type of situations. For example, the minimum description length (MDL) performs well for "small" models and poorly for "large" models while the Akaike's information criterion (AIC) does the opposite.

A.3.1 Akaike's Information Criterion

Akaike's Information Criterion (AIC) [101] was designed to be an approximately unbiased estimation of the Kullback-Leibler (KL) information divergence [122], a useful measure of the discrepancy between the true (unknown) model, f, and the fitted (approximating) model, g

$$\begin{aligned} KL(f,g) &= \int f(\boldsymbol{y}) \log\left(\frac{f(\boldsymbol{y})}{g(\boldsymbol{y}|\boldsymbol{\theta}_{\psi})}\right) d\boldsymbol{y} \\ &= \int f(\boldsymbol{y}) \log\left(f(\boldsymbol{y})\right) d\boldsymbol{y} - \int f(\boldsymbol{y}) \log\left(g(\boldsymbol{y}|\boldsymbol{\theta}_{\psi})\right) d\boldsymbol{y}, \end{aligned} \tag{A.6}$$

where the first term in Eq. (A.6) is equal for all models and can be neglected. Therefore, the AIC is an estimation of the expectation of the second term, multiplied by two. The model, which minimizes this term is to be chosen. The formula of the AIC can be expressed as

$$\mathrm{AIC}\left(\psi\right) = -2 \cdot l\left(\boldsymbol{\theta}_\psi|\boldsymbol{y}\right) + 2k_\psi, \tag{A.7}$$

where $l\left(\boldsymbol{\theta}_\psi|\boldsymbol{y}\right) = \log\left(g(\boldsymbol{y}|\boldsymbol{\theta}_\psi)\right)$ is the log-likelihood function of the model. The parameters vector $\boldsymbol{\theta}_\psi$ is estimated in the maximum likelihood sense and tends to be a minimizer of the KL divergence.

In the case of Gaussian distributed data, when the variance σ_ψ^2 has also to be estimated, a simplified formula of the AIC can be expressed as

$$\mathrm{AIC}\left(\psi\right) = N \log \hat{\sigma}_\psi^2 + 2k_\psi. \tag{A.8}$$

A.3.2 Corrected Akaike's Information Criterion

As the ratio k_ψ/N increases, AIC tends to be a strongly negatively biased estimation of the Kullback-Leibler information divergence, which consequently leads to overestimation. Corrected AIC (denoted $\mathrm{AIC_C}$) [102] is derived to correct for this bias. In [123], an approximate rule is stated: $\mathrm{AIC_C}$ should be used when $N/K_\psi < 40$, where K_ψ denotes the number of parameters in the model of the highest possible model-order, K . For linear models, $\mathrm{AIC_C}$ is exactly unbiased, assuming that the candidate models include the true one. In fact, $\mathrm{AIC_C}$ can be written as summation of AIC and a non-stochastic term, which means that the achieved reduction in bias does not increase the variance.

$\mathrm{AIC_C}$ can be expressed as [102]

$$\mathrm{AIC_C}\left(\psi\right) = N \log \hat{\sigma}_\psi^2 + N \frac{1 + k_\psi/N}{1 - (k_\psi + 2)/N}. \tag{A.9}$$

A.3.3 Minimum Description Length

Minimum Description Length (MDL) [103] adopts a different approach, than AIC and $\mathrm{AIC_C}$, to modeling. It proposes that the basic goal is to find regularities in data. Therefore, it is based on the idea that the proper way to capture regularities in data is to build a model which allows the shortest description of the data and the model itself. Hence, the motivation of the MDL is to measure the complexity of models and then select the least complex candidate model. The MDL searches for the model that minimizes the sum of the description (i.e., code) length of the model, plus the description (i.e., code) length of the data when fitted to this model.

From coding theory, a code, in which objects can be encoded, defines the probability distribution of the objects. Conversely, the probability distribution defines a code such that the code length of an object is $-\log_2 P(a)$ where $P(a)$ is the probability of the object. Therefore, minimizing the description length (i.e., code length) corresponds to maximizing the probability.

Consequently, MDL can be seen as a maximum likelihood principle (similar to AIC) and an approximation of the code length (which is used as a penalty term, stronger than the AIC one).

In case of linear Gaussian models, MDL can be written as

$$\mathrm{MDL}\,(\psi) = N \log \hat{\sigma}_{\psi}^2 + k_{\psi} \log N. \tag{A.10}$$

A.3.4 Normalized Maximum Likelihood

The Normalized Maximum Likelihood (NML) [124] is designed to be as an optimal universal distribution (a universal distribution $p_U(\boldsymbol{y})$, relative to a family of distributions M; each takes the form $f(\boldsymbol{y}|\hat{\boldsymbol{\theta}})$, is defined as a distribution that compresses each data set $\boldsymbol{y}$ almost as good as its maximum likelihood code (i.e., $-\ln f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{y}})$). There is an excess code length needed to encode the data set using this universal distribution, $\{-\ln p_U(\boldsymbol{y}) + \ln f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{y}})\}$, which is called the regret of $p_U(\boldsymbol{y})$ relative to M for the data.

Without making any assumptions to the unknown true distribution (exactly like the MDL approach), the optimal universal distribution can be formulated as a minimax problem: find the coding scheme which minimizes the worst-case expected regret [124]

$$p_1^*(\boldsymbol{y}) = \min_{p_1} \max_{p_2} E_{p_2}\left[\ln \frac{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{y}})}{p_U(\boldsymbol{y})}\right], \tag{A.11}$$

where p_2 ranges over the set of all probability distributions satisfying

$$E_{p_2}\left[\ln \frac{p_2(\boldsymbol{y})}{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{y}})}\right] < \infty. \tag{A.12}$$

The normalized maximum likelihood (NML) distribution is the distribution that satisfies this minimax problem, which can be formulated as

$$p_1^*(\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{y}})}{\int f(\boldsymbol{\tau}|\hat{\boldsymbol{\theta}}_{\boldsymbol{\tau}}) d\boldsymbol{\tau}}, \tag{A.13}$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\tau}}$ denotes the maximum likelihood estimation for the data set $\boldsymbol{\tau}$. The denominator of Eq. (A.13) is a normalizing constant (for which the name "normalized maximum likelihood" stands), which is the sum of maximum likelihoods of all data sets that could be found in an experiment.

The code length of the NML, $-\ln p_1^*(\boldsymbol{y}) = -\ln f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{y}}) + \ln \int f(\boldsymbol{\tau}|\hat{\boldsymbol{\theta}}_{\boldsymbol{\tau}})$, is called the stochastic complexity of the data set $\boldsymbol{y}$ with respect of the model class M, in which the second term is defined as the complexity of the model. The NML can be considered as an optimal implementation of the MDL principle.

As a basis for the denoising problem, which is an example of the linear regression problem, noise is defined as the part of the data that cannot be compressed with the considered models, while the compressible part defines the meaningful information-bearing signal. Inspired by that, the NML criterion has been derived for the case of general parametric model classes [125] and for the Gaussian distribution of noise [105].

For linear regression models with Gaussian distributed noise, the NML-based criterion in Eq. (A.13) can be manipulated and simplified to yield [105]

$$\text{NML}\,(\psi) = (N-k_\psi)\log \hat{\sigma_\psi}^2 + k_\psi \log(N\hat{R}) + (N-k_\psi-1)\log\left(\frac{N}{N-k_\psi}\right) - (k_\psi+1)\log k_\psi, \tag{A.14}$$

where

$$\hat{R} = \frac{1}{N}\hat{\boldsymbol{\theta}}_\psi^T \boldsymbol{X}_\psi^T \boldsymbol{X}_\psi \hat{\boldsymbol{\theta}}_\psi. \tag{A.15}$$

A.3.5 The Hannan-Quinn Criterion

Consistency and efficiency cannot occur mutually in a model selection criterion [126]. Basically, AIC is not strongly consistent (its penalty term does not depend on the sample size. Hence, there is a probability of overestimation), but efficient (it is able to select the true model-order with a probability approaching one as the sample size tends to be infinity and behaves "almost as well" in terms of mean squared error).

The Hannan-and-Quinn Criterion (HQC) [104] was proposed as an alternative to the AIC to provide a strongly consistent (almost surely) estimation of the correct model-order. It was shown in [104], that the HQC underestimates the true order to a less degree than the AIC for autoregressive models of large N or large regression coefficients.

Using the law of the iterated logarithm [126] to ensure strong consistency[4] of selection leads to the formula of Hannan and Quinn in the linear regression models

$$\mathrm{HQC}\left(\psi\right) = N \log \hat{\sigma_{\psi}}^2 + 2ck_{\psi} \log\log N, \tag{A.16}$$

where $c > 1$ is a threshold value. In [104], there was no advice of how to choose c.

A.3.6 Conditional Model-Order Estimator

Conditional model-order estimator (CME) [106] has a slightly different structure in terms of estimating the order based on the part of the data that does not depend on the model parameters. This can be achieved by transforming the observations (i.e., data) into a sufficient statistic[5] and an ancillary statistic[6] with respect to the model parameters. Then, the probability density function of the ancillary statistic [113] is used to estimate the model-order because the PDF of the ancillary statistic does not depend on the the model parameters but only on its dimension.

The CME was proposed to overcome the limitations of the Bayesian approach [119] (i.e., by avoiding any assumptions about the prior PDF of the model parameters) and those of the asymptotic criteria. It was shown in [106] that it works well for finite data records in terms of maximizing the probability of correct decision. It is applicable once sufficient statistics for the parameters of each competing model are existed.

In [106], it was mentioned that the CME takes the form of the maximum likelihood (ML) rule[7]. There, the CME was derived for model-order selection of the Gaussian linear model of Eq. (A.1) with unknown variance, σ^2. For such a model, the unknown set of parameters, $\left[\boldsymbol{\theta}\sigma^2\right]^T$, have the PDFs

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}\left(\hat{\boldsymbol{\theta}}, \sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right), \tag{A.17}$$

and

$$u = \frac{\left(N-k\right)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-k}, \tag{A.18}$$

---

[4] There are two types of consistency, weak and strong. A model selection method is weakly consistent if, with probability approaches one as the sample size tends to infinity, the selection method is able to select the true (or the closest-to-true in some sense) model from the candidate models. Strong consistency is attained when the selection of the true model (or the closest-to-true in some sense) is almost surely (i.e., happens with probability one).

[5] A statistic T is a sufficient statistic for a parameter θ when the conditional probability density function of the data given the statistic does not depend on θ.

[6] A statistic T is an ancillary statistic for a parameter θ if the probability density function of T does not depend on θ.

[7] Therefore, we assign it here to the category of information theoretic-based criteria.

respectively. Then, a sufficient statistic for the model parameters from the data, $\boldsymbol{T}(\boldsymbol{y})$, was derived and thus the likelihood function was calculated as follows

$$L_{Y|T}(\boldsymbol{y}) = \frac{p_Y(\boldsymbol{y};\boldsymbol{\theta},\sigma^2)}{p_T(\boldsymbol{T}(\boldsymbol{y});\boldsymbol{\theta},\sigma^2)} \tag{A.19}$$

After some manipulations, the CME for model and model-order selection can be expressed as [106]

$$\mathrm{CME}(\psi) = \frac{N-k_\psi-2}{2}\log\hat{\sigma_\psi}^2 + \frac{1}{2}\ln|\boldsymbol{X}_\psi^T\boldsymbol{X}_\psi| + \log\left(\frac{[\pi(N-k_\psi)]^{(N-k_\psi)/2}}{\Gamma\left(\frac{N-k_\psi}{2}\right)}\right), \tag{A.20}$$

where $\Gamma(\cdot)$ is Euler's Gamma function.

A.3.7 Mallows' $\mathrm{C_p}$ Criterion

Generally, the quality of the estimated (selected) model can be objectively measured as follows

$$\boldsymbol{J}_\psi = \frac{1}{\sigma^2}\left(\boldsymbol{X}_\psi\hat{\boldsymbol{\theta}}_\psi - \boldsymbol{X}\boldsymbol{\theta}\right)^T\left(\boldsymbol{X}_\psi\hat{\boldsymbol{\theta}}_\psi - \boldsymbol{X}\boldsymbol{\theta}\right), \tag{A.21}$$

which is a distance measure between the estimated model and the true noise-free model, $\boldsymbol{X}\boldsymbol{\theta}$. It is clear that this measure is only possible in simulations as the true model has to be known.

For practical applications, the $\mathrm{C_p}$ criterion is used because it has the same expectation as $\boldsymbol{J}_\psi$. The $\mathrm{C_p}$-plot is a graphical display that is helpful to analyze the data at hands [107]. In such a way of analysis, adequacy of prediction can be measured in terms of examining the influence of including different sub-models on the behaviour of residuals sum of squares.

In its basic form, the Mallows' $\mathrm{C_p}$ criterion [107] has a behaviour similar to that of the AIC. It can be expressed as

$$\mathrm{C_p}(\psi) = \frac{\mathrm{RSS}_{k_\psi}}{\hat{\sigma}^2} - N + 2k_\psi, \tag{A.22}$$

where RSS_{k_ψ} is the residuals sum of squares for the model ψ that is computed as follows

$$\mathrm{RSS}_{k_\psi} = \left(\boldsymbol{y} - \boldsymbol{X}_\psi\hat{\boldsymbol{\theta}}_\psi\right)^T\left(\boldsymbol{y} - \boldsymbol{X}_\psi\hat{\boldsymbol{\theta}}_\psi\right). \tag{A.23}$$

Here, $\hat{\sigma}^2$ is the variance computed in the largest model when all regression parameters are included.

A generalized version of the Mallows' $\mathrm{C_p}$ criterion was proposed in [127] as follows

$$\mathrm{C_p}\left(\psi, \beta\right) = \frac{\mathrm{RSS}_{k_\psi}}{\hat{\sigma}^2} - N + \beta k_\psi, \tag{A.24}$$

where β is a penalty factor. There is, however, no optimal β value for all possible situations [127]. It was shown in [128] that for moderate SNR values, $\beta = 3$ yielded a good balance between underestimation and overestimation for autoregressive models while in [127] the aim was to select β values that yield accurate models of true derivatives of the data. Since derivatives are more sensitive to overestimation (overfitting), it is recommended to avoid overestimation when it is required to obtain models with good derivatives' approximation ability. This can be fulfilled by increasing the penalty factor, β.

A.3.8 Efficient Detection Criterion

One of the interesting points in model and model-order selection problems is the proper choice of the penalty function of the used criterion to be suitable for the problem at hands. The efficient detection criterion (EDC) has the advantage over other information theoretic-based criteria in terms of its capability of using non-fixed (flexible) penalty functions.

In [108], the EDC was introduced for detecting the number of signals in the presence of Gaussian white noise under an additive model and its strong consistency was also proven.

In [109], the EDC was introduced in regression problems as

$$\mathrm{EDC}\left(\psi\right) = N \log \hat{\sigma}_\psi^2 + k_\psi C_N, \tag{A.25}$$

where C_N satisfies

$$\lim_{N \to \infty} \left\{ \frac{C_N}{N} \right\} = 0, \tag{A.26}$$

and

$$\lim_{N \to \infty} \left\{ \frac{C_N}{\log \log N} \right\} = \infty, \tag{A.27}$$

to ensure the condition of strong consistency.

There are many choices of C_N which ensure Eqs. (A.26) and (A.27). In [109], an *ad hoc* method was used to select C_N, which was in the form δN^γ where $\gamma < 1$. Different combinations of (δ, γ) were tried out and the values which gave the highest frequency of selecting the correct model were adopted for the given model. However, it was mentioned in [109] that further research is needed for the proper choice of C_N.

A.4 Resampling-Based Model Selection Criteria

Recently, resampling-based criteria were widely used in model and model-order selection. They showed superior performance to those achieved with information theoretic-based criteria in a range of simulated applications [11, 12, 117] and real life problems [6, 8, 13, 129].

Resampling-based criteria are used to approximate the resampling-based estimate of the mean-squared prediction error [130]. They select the model (or model-order) that minimizes the following "score function"

$$F(\psi) = \frac{1}{B} \sum_{b=1}^{B} SSE^{*}_{N,m}(\psi)_b, \tag{A.28}$$

where $SSE^{*}_{N,m}$ is the resampling-based sum of squared errors, B is the number of resamples, and m is a scaling factor or a replacement of it, as explained below.

In this Section, two resampling-based algorithms for model (model-order) selection are introduced and explained. The linear model in Eq. (A.1) holds here but the only difference is the possibility of relaxing the assumption of Gaussian noise. This means that $\boldsymbol{w}$ can be generally assumed as an $N \times 1$ iid noise vector of unknown distribution with mean zero and variance σ^2.

A.4.1 Bootstrap Based Model Selection

The steps of the procedure for using the boostrap for model and model-order selection are provided as follows [12]:

1. Given the data, $\boldsymbol{y}$, compute the LSE of the parameters of the largest model, $\psi = \psi_{\max}$ (see Eq. (A.3)), and compute an approximation of the data

$$\hat{\boldsymbol{y}} = \boldsymbol{X}_{\psi_{\max}} \hat{\boldsymbol{\theta}}_{\psi_{\max}}. \tag{A.29}$$

2. Compute the residuals: $\hat{\boldsymbol{r}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$.

3. Rescale and detrend the residuals

$$\tilde{\boldsymbol{r}} = s\sqrt{\frac{N}{m}} \left(\hat{\boldsymbol{r}} - \frac{1}{N} \sum_{n=1}^{N} \hat{r}_i \right), \tag{A.30}$$

where $s = 1$ for model-order selection and $s = 1/\sqrt{(1 - k_\psi)}$ for model selection, and m is the scaling parameter introduced such that $\lim_{N\to\infty} m/N = 0$ and $\lim_{N\to\infty} m = \infty$ [110]. The parameter m depends on the model parameters and is chosen in a way to satisfy enough variability amongst different bootstrap resamples. Accordingly, m needs always tuning.

4. For all models ψ:
   (a) Calculate $\hat{\boldsymbol{\theta}}_\psi$ as in Eq. (A.3) and $\hat{\boldsymbol{y}}$ as in Eq. (A.29).
   (b) Resample with replacement from the empirical distribution of $\tilde{\boldsymbol{r}}$ to obtain the bootstrap residuals $\tilde{\boldsymbol{r}}^*$.
   (c) Define a pseudo-data vector: $\boldsymbol{y}^* = \hat{\boldsymbol{y}} + \tilde{\boldsymbol{r}}^*$.
   (d) Compute $\hat{\boldsymbol{\theta}}^*_\psi$ and then $\hat{\boldsymbol{y}}^* = \boldsymbol{X}_\psi \hat{\boldsymbol{\theta}}^*_\psi$.
   (e) Compute the bootstrap sum of squared errors

$$SSE^*_{N,m}(\psi) = \frac{1}{N}\sum_{n=1}^{N} \left(\boldsymbol{y} - \hat{\boldsymbol{y}}^*\right)^2 . \tag{A.31}$$

   (f) Repeat steps (b)–(e) B times and then Compute the score function

$$F(\psi) = \frac{1}{B}\sum_{b=1}^{B} SSE^*_{N,m}(\psi)_b. \tag{A.32}$$

5. Choose ψ that yields the minimal value of $F(\psi)$.

A.4.2 Hook-and-Loop Based Model Selection

In practical cases of model selection, the choice of the scaling parameter m is not always obvious. In some cases this choice may become critical if one aims to show the superiority of the bootstrap methods to other information theoretic-based model selection criteria [6, 17–19]. A popular alternative to scaling is based on subsampling [98] in which rather than using full data record of N samples, subsamples of m are used to achieve the consistency in the estimation of the bootstrap sum of squared errors, $SSE^*_{N,m}(\psi)$. However, the choice of the subsample length m may also be a problem in particular when N is small.

In [20, 21], a new heuristic resampling scheme, the “Hook-and-Loop” (HL), was introduced for avoiding scaling problems. It uses the same concept of bootstrap in terms of resampling-with-replacement from the empirical distributions of the residuals. The core of the HL-based model and model-order selection is identical to the bootstrap-based one provided in Section A.4.1. The only difference lies in the steps of detrending the residuals and resampling them. We highlight here these difference as follows:

1. The residuals, $\hat{\boldsymbol{r}}$, are computed as in steps (1)–(2) in the boostrap procedure.
2. Rescale and detrend the residuals

$$\tilde{\boldsymbol{r}} = s\sqrt{N}\left(\hat{\boldsymbol{r}} - \frac{1}{N}\sum_{n=1}^{N}\hat{r}_i\right), \tag{A.33}$$

   where $s = 1$ for model-order selection and $s = 1/\sqrt{(1-k_\psi)}$ for model selection.
3. Sort the detrended residuals, $r_1, r_2, \ldots, r_N$, in an ascending order to obtain $r_{(1)} < r_{(2)} < \ldots < r_{(n)}$.
4. Generate a new HL sample using: $r^*_{(i)} = \frac{1}{2}\left(r_{(i)} + r_{(i+1)}\right) + \varphi_i$, where

$$\varphi_i \sim \mathcal{N}\left(0, \left[\frac{1}{6}\left(r_{(i+1)} - r_{(i)}\right)\right]^2\right). \tag{A.34}$$

5. Add one random point from the original residual sample to make the length of the HL sample equal to that of the original signal.
6. Re-order the HL residuals according to the strength of the original signal to obtain $\tilde{\boldsymbol{r}}^*$.
7. Follow on from step (4) in the bootstrap-based procedure.

The original and the HL sample form a kind of a zipper or a fastener, hence the name “hook-and-loop”. The idea of sorting the pseudo-new residuals has been introduced to match the strength of the HL residual to that of the original signal (i.e., small values together and large values together) to eliminate the need for scaling the residuals.

A.5 Biased Estimation in Linear Models

Rather than using the LSE method for estimating the parameters of the linear model of Eq. (A.1), there are other estimation methods with different properties that can be used in the framework of model and model-order selection. For example, biased estimation [131,132], Tikhonov regularization [133,134], and Least Absolute Shrinkage and Selection Operator (LASSO) [135].

In [121], it was shown that biased estimators are capable of outperforming MVU estimators in view of minimizing the MSE of an estimator. Biased estimation can be accomplished by scaling (i.e., shrinking) the unbiased estimator in order to get a biased estimator which has a smaller MSE than the unbiased one. It has been proved in [121] that for a wide range of theoretical models, such a scaling factor can be found using convex optimization techniques.

To implement the biased estimator, $\hat{\theta}_b$, it can be assumed as a scaled version of the unbiased one, $\hat{\theta}_u$, as given by [121]

$$\hat{\theta}_b = (1+\upsilon)\,\hat{\theta}_u, \tag{A.35}$$

where υ is the scaling factor. Then, the MSE of the biased estimator can be expressed as

$$\mathrm{MSE}(\hat{\theta}_b) = (1+\upsilon)^2\,\mathrm{var}(\hat{\theta}_u) + \upsilon^2\theta^2. \tag{A.36}$$

The parameter υ is to be chosen so that $\mathrm{MSE}(\hat{\theta}_b) < \mathrm{MSE}(\hat{\theta}_u)$ for all θ and such that the reduction in MSE is as large as possible. This can be formulated as solving a minimax problem [136,137]

$$\upsilon_o = \arg\min_{\upsilon}\max_{\theta}\{\mathrm{MSE}(\hat{\theta}_b) - \mathrm{MSE}(\hat{\theta}_u)\}. \tag{A.37}$$

An extension of Eq. (A.37) to multiple parameters can be expressed as [121]

$$\boldsymbol{\Upsilon}_o = \arg\min_{\Upsilon}\max_{\theta\in\mathcal{U}}\{\mathrm{MSE}(\hat{\boldsymbol{\theta}}_b) - \mathrm{MSE}(\hat{\boldsymbol{\theta}}_u)\}. \tag{A.38}$$

In this work, we investigate the applicability of using biased estimation in the context of model and model-order selection [19]. This is done by scaling the LSE of parameters vector, $\hat{\boldsymbol{\theta}}_\psi$, that is considered as $\hat{\boldsymbol{\theta}}_u$ to obtain $\hat{\boldsymbol{\theta}}_b$, according to Eq. (A.35).

A.6 Model-Order Selection by Partial-Model Selection Search

The traditional searching method for model-order selection in linear regression is a nested full-parameters-set searching procedure over the desired orders, $k \in \{1, 2, \ldots, K\}$, which we call full-model order selection. Finding the optimal order can thus be expressed as

$$\hat{k}_o = \arg \min_{k \in \{1,2,\ldots,K\}} \{F(k)\}, \tag{A.39}$$

where $F(k)$ is a criterion score[8], either information theoretic- or resampling-based.

On the other hand, the method of model selection searches for the best sub-model, $\psi \in \{\psi_1, \psi_2, \ldots, \psi_{2^K}\}$, which can be expressed as

$$\hat{\psi}_o = \arg \min_{\psi \in \{\psi_1,\psi_2,\ldots,\psi_{2^K}\}} \{F(\psi)\}, \tag{A.40}$$

where we consider here that the maximum number of sub-models is 2^K.

We propose using the model selection searching method for model-order selection, which we call partial-model order selection [18]. The proposed searching method, which includes basically the best sub-model estimation within each model-order, is outlined as follows:

1. For each model-order $k = 1, 2, \ldots, K$:
   (a) Construct the corresponding sub-models. There are 2^{k-1} sub-modes that maintain the parameter of the kth order

   $$\{\psi_{k_{(i)}} : i = 1, 2, \ldots, 2^{k-1}, \quad \text{with } \theta_k \neq 0\}, \tag{A.41}$$

   $$\boldsymbol{y} = \boldsymbol{X}_{\psi_k} \boldsymbol{\theta}_{\psi_k} + \boldsymbol{w}. \tag{A.42}$$

   (b) Compute the LSE of $\boldsymbol{\theta}_{\psi_k}$

   $$\hat{\boldsymbol{\theta}}_{\psi_k} = \left(\boldsymbol{X}_{\psi_k}{}^T \boldsymbol{X}_{\psi_k}\right)^{-1} \boldsymbol{X}_{\psi_k}{}^T \boldsymbol{y}. \tag{A.43}$$

   (c) Compute a vector of criterion scores (information theoretic- or resampling-based)

   $$\boldsymbol{F}_{\psi_k} = \{F(\psi_{k_{(i)}}) : i = 1, 2, \ldots, 2^{k-1}\}. \tag{A.44}$$

---

[8] Note that the model we search for is the order. This is the reason why $F(\psi)$ was replaced by $F(k)$.

(d) Select a representative score for order k

$$F(k) = \min_{\psi_k}(\boldsymbol{F}_{\psi_k}). \tag{A.45}$$

2. Estimate the optimal order

$$\hat{k}_o = \arg \min_{k \in \{1,2,\ldots,K\}} \{F(k)\}. \tag{A.46}$$

From the steps of the proposed algorithm, we observe that there are two steps of minimization. The first step searches for a representative score for the current model-order, k, we try to fit our data to. Then, in the second minimization step, the optimal model-order, $\hat{k}_o$, is estimated by taking the minimum of all representative scores.

A.7 Simulation Results with Discussion

To illustrate the different concepts presented above, simulation results for biased versus unbiased estimation in the context of model and model-order selection are presented. Moreover, a comparison between full- and partial-model order selection searching methods for different model selection criteria are provided.

A.7.1 Biased versus Unbiased Estimation

We present the simulation results for two examples in the context of model and model-order selection. We compare the performance of some of the different approaches described above across different Signal-to-Noise Ratios (SNR) for both biased and unbiased estimators. In all simulations, the bootstrap and the HL repetitions are set to $B = 100$ and the scaling parameter, m, is set to 2 for the bootstrap procedure [12,20,21]. Only Gaussian noise models with iid random variables are considered here. It was shown in [10] that bootstrap techniques perform better than information criteria for model selection in the non-Gaussian case. It is not the objective to demonstrate here that the bootstrap is superior also when using biased estimators. In this work, we are primarily concerned with the effect of biased estimation on some model selectors.

A.7.1.1 Model Selection

We choose a polynomial from [12] given by

$$y = 0 + 0 + 0.035x^2 - 0.0005x^3 + w, \quad x = 0, \ldots, N-1, \tag{A.47}$$

to evaluate the performance of biased (B) and unbiased (U) estimators for model selection. The model is then normalized over N. Table A.1 shows the empirical probabilities of selecting the correct model using the MDL, the bootstrap and the HL. Different signal lengths ($N = 32$, $N = 64$ and $N = 128$) were used with an SNR ranging between 0 dB and 40 dB.

Table A.1. Empirical probabilities (in %) of selecting the correct model evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (A.47) with Gaussian noise. The highest probabilities achieved for each of the cases are indicated in bold fonts.

N	SNR [dB]	MDL		Bootstrap		HL	
		U	B	U	B	U	B
32	0	0.5	0.5	0.0	0.0	**1.0**	**1.0**
	10	14.7	11.5	0.5	0.4	**52.1**	45.0
	20	87.1	88.5	74.9	75.1	**95.0**	28.0
	30	95.5	97.1	99.5	**99.7**	97.4	93.5
	40	95.9	97.7	99.7	**99.8**	97.8	94.6
64	0	39.4	39.1	0.5	0.3	**47.4**	38.7
	10	96.2	96.2	74.3	73.9	**97.2**	96.1
	20	96.7	97.9	**99.7**	**99.7**	**99.7**	98.7
	30	98.1	99.4	99.9	**100.0**	**100.0**	98.8
	40	98.1	99.2	99.7	**99.9**	99.8	98.8
128	0	46.5	48.0	0.1	0.3	**50.5**	18.4
	10	90.6	**91.2**	63.1	61.9	82.3	77.2
	20	98.8	99.1	99.5	99.4	**100.0**	**100.0**
	30	98.9	99.3	99.9	**100.0**	**100.0**	**100.0**
	40	98.3	99.0	99.6	99.6	**100.0**	**100.0**

We observe from Table A.1 that for an SNR larger than 20 dB, the biased estimators yield higher probabilities of selecting the correct model than the unbiased ones for the MDL criterion, but not necessarily for the bootstrap and HL model selection schemes. For the bootstrap, the application of a biased estimator increases sometimes the probability of selecting the correct model, but it does not for the HL method when the probability of correct detection decreases, except when both the SNR and the sample size are high. It is worth noting that a biased estimator for bootstrap model selection leads to higher probabilities as compared to the biased ones for the MDL criterion. In Table A.1, the highest probabilities achieved for each of the cases are indicated in bold fonts, showing that the overall performance achieved with the HL method based on unbiased estimators is superior to the MDL and the bootstrap based technique counterparts.

A.7.1.2 Model-Order Selection: Example 1

For evaluating the performance of the biased estimators in the context of non-iterative model-order selection (i.e., the MDL and CME criteria), we choose the polynomial trend from [106] in noise, given by

$$y = 0 + 0.4x + 0.1x^2 + w, \quad x = 0, \ldots, N-1, \tag{A.48}$$

In Table A.2, we compare the empirical probabilities of selecting the correct model-order (second order in this case) for a signal length of $N = 30$ and a noise variance, ranging from $\sigma^2 = 10$ to $\sigma^2 = 50$. The highest possible model-order was set to 10.

Table A.2. Empirical probabilities (in %) of correct selection of the model-order evaluated over 1000 independent Monte Carlo runs for the model in Eq. (A.48) with Gaussian noise. The highest probabilities achieved for each of the cases are indicated in bold fonts.

	MDL		CME	
σ^2	U	B	U	B
10	94.0	97.7	**100.0**	**100.0**
15	92.5	94.4	**100.0**	**100.0**
20	93.2	92.6	**100.0**	99.9
25	94.3	90.9	**100.0**	99.9
30	92.3	89.1	**99.9**	99.6
35	92.9	86.8	**99.8**	98.4
40	93.8	86.3	**99.7**	96.2
45	93.6	84.1	**99.3**	93.5
50	92.4	78.1	**99.3**	87.9

We observe from Table A.2 that the unbiased estimators lead to a CME which performs better than the MDL (as proved in [106]). Moreover, biased estimators for the CME yield higher accuracies than their counterparts of the MDL for all noise variances. However, biased estimators give lower probabilities than unbiased ones for both the CME and the MDL criteria.

A.7.1.3 Model-Order Selection: Example 2

We use the polynomial of Eq. (A.47) to evaluate the performance of biased and unbiased estimators for model-order selection (third order in this case) using the MDL, the CME, the bootstrap and the HL. Different signal lengths ($N = 32$, $N = 64$ and $N = 128$) were used with an SNR, ranging between 0 dB and 40 dB, as shown in Table A.3. The highest considered model-order was set to 10.

Table A.3. Empirical probabilities (in %) of selecting the correct model-order evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (A.47) with Gaussian noise. The highest probabilities achieved for each of the cases are indicated in bold fonts.

	SNR	MDL		CME		Bootstrap		HL	
N	[dB]	U	B	U	B	U	B	U	B
32	0	**0.9**	0.8	0.0	0.0	0.0	0.0	0.0	0.0
	10	**3.3**	2.1	0.0	0.0	0.2	0.0	0.0	0.5
	20	**35.9**	24.2	0.0	3.5	13.5	6.6	29.0	32.0
	30	**96.4**	79.1	0.0	47.8	95.2	38.1	84.5	44.4
	40	93.8	97.5	0.0	91.1	**97.8**	96.1	91.9	82.2
64	0	**3.9**	3.2	0.0	0.0	0.0	0.0	0.0	0.0
	10	**88.9**	41.7	0.0	1.6	12.2	4.7	0.0	0.0
	20	96.6	79.7	0.0	29.4	**98.3**	73.2	11.7	12.5
	30	98.5	97.0	0.0	94.0	**99.2**	95.4	91.6	70.5
	40	97.2	**99.5**	0.0	98.7	98.7	99.3	89.9	60.2
128	0	**18.7**	12.5	0.0	0.1	0.0	0.1	0.0	0.0
	10	**98.0**	41.0	0.0	2.2	8.7	4.8	0.0	0.0
	20	98.0	85.6	0.0	46.5	**98.6**	73.9	60.0	10.7
	30	98.3	99.2	0.0	89.4	99.3	99.7	**100.0**	66.1
	40	98.6	99.2	0.0	97.2	98.5	99.6	**100.0**	77.0

The example chosen here shows a breakdown of the CME, which always selects the highest possible model-order (i.e., 10 rather than 3). Interestingly, the CME regains some of its power when a biased estimator is used, but its performance is still lower than that of the corresponding MDL. This indicates that biased estimates for the CME lead to better results than unbiased estimates. The results also indicate that biased estimators can lead to a slightly higher probability of selecting the correct model-order at high SNR, but not for the HL method. Overall the performance of the unbiased estimator based MDL is best for a low SNR while the bootstrap and HL excel at a higher SNR values.

A.7.2 Partial- versus Full-Model Order Selection Searching Methods

We present simulation results for four examples of different complexities. We compare the accuracies of selecting the correct model-order for the different criteria described above across different SNR values for both the traditional (the full-model order, denoted by “F” in the results) and the proposed (the partial-model order, denoted by “P” in the results) searching methods. In all simulations, the following settings are used: 1) sample size of $N = 30$, 2) SNR ranges from 5 to 20 dB, 3) the highest possible model order is $K = 10$, 4) 1000 independent Monte Carlo runs, 5) Gaussian noise models with independent and identically (i.i.d) distributed random variables, 6) $c = 1.5$ for the HQC, 7) 100 bootstrap replications, 8) $m = 2.975$ for the bootstrap procedure, and 9) $G(k, N) = kN^{0.5}$ for the EDC. It is worthy to mention that the value $m = 2.975$ was selected by performing an independent on-the-fly study by taking the value in [12] as an initial guess and then fine-tune it.

A.7.2.1 Example 1

We choose a polynomial trend from [106] in noise, given by

$$y = 0 + 0.4x + 0.1x^2 + w, \quad x = 0, \ldots, N-1. \tag{A.49}$$

In Table A.4, we compare the empirical probabilities of selecting the correct model-order (second order in this case).

From Table A.4, we observe that the proposed search method outperforms the traditional one for all used criteria and all SNR values. Interestingly, using the proposed method with the bootstrap shows significant accuracy improvement at low SNR, 5 and 10 dB. However, for high SNR the performance of the two searching methods is relatively comparable for all criteria, except for the bootstrap which shows better performance when the full-model order selection method is used. This polynomial trend was used in [106] to show the outperformance of the CME criterion over the MDL one. However, we observe that the CME shows outstanding performance only for high SNR, while the bootstrap outperforms it for low SNR.

Table A.4. Empirical probabilities (in %) of selecting the correct model-order evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (A.49) with Gaussian noise. The highest frequencies achieved for each column are indicated in bold fonts.

Criterion	SNR [dB] 5 P	5 F	10 P	10 F	15 P	15 F	20 P	20 F
MDL	37.7	25.7	72.9	64.1	91.7	91.4	94.3	93.9
AIC	48.6	**34.6**	73.0	66.7	81.3	81.7	80.6	83.4
$\mathrm{AIC_C}$	45.5	30.9	75.8	**67.7**	90.0	89.1	91.3	92.3
HQC	34.2	24.2	71.7	62.3	92.6	92.3	95.1	94.5
CME	46.6	30.7	70.1	62.2	**96.8**	**96.7**	**100.0**	**100.0**
NML	37.8	24.8	66.0	57.6	95.0	94.0	99.3	99.1
BOOT	**60.9**	16.6	**81.6**	50.2	79.4	89.5	70.9	95.7
EDC	20.4	14.6	60.8	50.9	94.2	93.3	98.0	98.0

A.7.2.2 Example 2

For further evaluation of the performance of the proposed method, we use another second-order polynomial trend which contains full non-zero parameters set

$$y = 0.8 + 0.5x + 0.3x^2 + w, \quad x = 0, \ldots, N-1. \tag{A.50}$$

This polynomial is relatively easier than the one of Eq. (A.49) for model-order selection criteria to work on. However, it has full non-zero parameter set which means that the full-model should be the correct one. In Table A.5, we show that this is not always the case and the partial-model order selection method performs better than the full-model one, significantly at low SNR.

A.7.2.3 Example 3

We use a third-order polynomial trend from [12], which is of a higher complexity than the ones in Eq. (A.49) and Eq. (A.50)

$$y = 0 + 0 + 0.035x^2 - 0.0005x^3 + w, \quad x = 0, \ldots, N-1. \tag{A.51}$$

This model is complex in the sense that the contribution from the coefficient of the third order is very small in comparison to the second order one. Hence, model-order selection criteria perform poorly for such a model (except for very large SNR) and usually underestimate the correct order.

Table A.5. Empirical probabilities (in %) of selecting the correct model-order evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (A.50) with Gaussian noise. The highest frequencies achieved for each column are indicated in bold fonts.

Crit-	SNR [dB]							
	5		10		15		20	
erion	P	F	P	F	P	F	P	F
MDL	41.6	28.4	76.4	69.5	94.3	94.1	93.3	93.4
AIC	50.0	34.9	75.1	69.5	83.8	83.5	82.4	82.3
$\mathrm{AIC_C}$	47.5	33.5	78.3	72.2	92.4	92.5	91.5	91.0
HQC	38.9	26.9	75.9	67.6	95.6	95.0	94.4	94.2
CME	**76.4**	**63.4**	**89.5**	**86.5**	**99.6**	**99.5**	**99.8**	**99.6**
NML	41.2	28.1	72.4	63.1	98.1	97.7	98.8	98.6
BOOT	61.7	18.6	82.9	57.8	87.6	94.6	84.1	95.1
EDC	24.0	17.2	66.5	57.4	98.0	97.5	98.0	97.8

In Table A.6, we compare the performance of the two selection methods for the correct (third) order and for the second order. Using the proposed searching method gives higher accuracies for selecting the realistic order for such a polynomial model (second order in this case) for low SNR (5- and 10 dB) and all used selection criteria. On the other hand, for higher SNR (15- and 20 dB), the criteria start working and the performance of the partial-model selection method outperforms the full-model one, for almost all selection criteria. However, the CME is an exception as it completely fails for this example.

A.7.2.4 Example 4

We choose a less-complex third-order polynomial trend in noise

$$y = 1 + 0.5x + 0.03x^2 - 0.05x^3 + w, \quad x = 0, \ldots, N-1. \tag{A.52}$$

Table A.7 shows the empirical probabilities of selecting the correct (third) order. It is clear that the proposed order selection method provides better accuracies than the traditional one for all model-order selection criteria and across all SNR values. Moreover, combining the proposed searching method with the bootstrap-based criterion significantly outperforms all other criteria for all SNR values.

Table A.6. Empirical probabilities (in %) of selecting the third order and the second order evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (A.51) with Gaussian noise. The highest frequencies for selecting the second order (the third order) achieved for each column of the 5- and 10 dB (the 15- and 20 dB) are indicated in bold fonts.

Criterion	SNR [dB]							
	5		10		15		20	
	P	F	P	F	P	F	P	F
MDL_3	0.9	1.2	3.8	4.9	12.6	11.1	42.8	29.7
MDL_2	23.4	15.5	44.5	35.5	68.8	69.2	53.8	66.6
AIC_3	3.5	4.8	7.0	8.4	**22.5**	**18.7**	**52.2**	**37.4**
AIC_2	33.9	**21.7**	49.7	**42.5**	57.8	62.1	35.8	48.9
AIC_{c3}	2.2	2.4	4.9	6.4	16.4	14.0	48.3	33.8
AIC_{c2}	29.6	19.9	48.7	40.8	68.1	69.5	47.3	61.8
HQC_3	0.7	0.8	3.3	4.2	10.8	10.0	39.3	28.1
HQC_2	21.4	14.9	43.0	34.1	69.7	69.0	57.9	68.9
CME_3	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.7
CME_2	6.8	5.6	15.5	10.7	45.8	40.9	93.4	93.3
NML_3	0.7	0.9	2.1	2.3	3.1	3.6	14.4	11.9
NML_2	23.1	15.3	35.5	27.7	66.5	63.5	84.5	87.0
$BOOT_3$	6.7	0.4	3.7	2.8	16.6	6.3	34.5	21.3
$BOOT_2$	**51.9**	10.5	**58.0**	23.6	36.9	62.6	12.3	76.2
EDC_3	0.3	0.3	0.9	1.4	4.4	4.7	23.4	19.7
EDC_2	11.5	8.5	30.4	21.6	67.1	64.1	75.4	78.9

A.8 Summary

Model and model-order selection are well-known and well-researched problems. There are variety of criteria both information theoretic- and resampling-based that can be utilized for solving such problems. However, it is always difficult to draw a conclusion about the superiority of a single criterion as some criteria are more suitable for some applications while they can fail for other applications. In this Appendix, a wide range of different model and model-order selection criteria for linear regression polynomials have been provided.

Besides that using unbiased estimation of the parameters, the applicability of biased estimation in model and model order selection has been investigated where further minimization of the MSE via biased estimation could lead to better results. However, there were no theoretical results for the consistency for model and model-order selection with biased estimators as we primarily were interested in evaluating the behavior of a

Table A.7. Empirical probabilities (in %) of selecting the correct model-order evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (A.52) with Gaussian noise. The highest frequencies achieved for column are indicated in bold fonts.

Criterion	SNR [dB]							
	5		10		15		20	
	P	F	P	F	P	F	P	F
MDL	7.1	6.4	21.2	15.7	54.9	34.5	85.0	78.7
AIC	17.0	**12.5**	36.4	**19.8**	63.9	**43.9**	81.5	76.6
AIC$_\mathrm{C}$	9.9	7.7	27.5	17.1	60.1	39.8	**87.3**	**81.2**
HQC	6.3	5.7	18.6	14.5	51.9	32.9	84.1	77.1
CME	16.7	12.3	19.5	14.8	38.9	25.5	73.7	61.1
NML	6.6	6.2	12.2	10.5	30.7	21.8	69.4	55.4
BOOT	**42.9**	4.3	**67.6**	9.3	**84.5**	23.9	**87.2**	64.6
EDC	2.9	3.2	9.1	8.2	33.5	22.9	77.5	67.6

few model selection techniques in view of biased estimation. Techniques such as the AIC, the corrected AIC, and the HQC yielded very similar results to those of the MDL and hence we restricted our choice here to the MDL criterion. Emphasis was given to modern techniques such as the CME, the bootstrap and the HL model selectors. We have shown that model selection with biased estimators can sometimes slightly increase the probability of selecting the correct model but this is not true for model order selection, except for the CME.

To use the advantage of sparsity in model selection based criteria, we have proposed a partial-model selection search technique to improve the performance of the model-order selection criteria. The improvement of the model-order selection criteria becomes evident for lower SNR values. On average (mean ± SD), for all four models and all eight model selection criteria that were considered, the improvement in selecting the correct model order was $(17 \pm 14)\%$ and $(68 \pm 70)\%$, for SNR of 10 dB and 5 dB, respectively. For SNR $=$ 5 dB, the bootstrap-based method achieved the highest improvement of over 300% while this was not so evident for higher SNR values. It is worth noting, that the performance of the bootstrap method depends on the scaling parameter m, chosen here to be $m = 2.975$, which can be individually tuned to the SNR level.

List of Acronyms

AIC	Akaike's Information Criterion
AICc	Corrected Akaike's Information Criterion
BOOT	Bootstrap
CCD	Charged-Couple Devices
CDF	Cumulative Distribution Function
CME	Conditional Model-Order Estimation
Cp	Mallow's Cp Criterion
D	Dioptres
E	Eccentricity
EDC	Efficient Detection Criterion
FAR	False Alarm Rate
HL	Hook-and-Loop Resampling Plane
HQC	Hannan-Quinn Information Criterion
HSS	Hartmann-Shack wavefront Sensor
iid	independent and identically distributed
KL	Kullback-Leibler
KW	Kruskal-Wallis Test
LASSO	Least Absolute Shrinkage and Selector Operator
LASIK	Laser-Assisted in SItu Keratomileusis
LSE	Least Square Estimation
MAP	Maximum *A Posteriori*
MDL	Minimum Description Length
minK	Corneal power in the flattest corneal meridian
ML	Maximum Likelihood

MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MVU	Minimum Variance Unbiased estimator
NML	Normalized Maximum Likelihood
PDF	Probability Density Function
RMS	Root Mean Square
RSS	Residuals Sum of Squares
SAI	Surface Asymmetry Index
SE	Structure Element
SNR	Signal-to-Noise Ratio
simK	Simulated Keratometric Values
SRI	Surface Regularity Index
3D	Three-Dimensional
2D	Two-Dimensional
VKS	Videokeratoscopic/Videokeratoscopy

List of Symbols

$\mathbf{0}$	Vector of zeros
$\underset{x}{\mathrm{argmax}}\, y$	Returns the value of x that maximizes y
$\underset{x}{\mathrm{argmin}}\, y$	Returns the value of x that minimizes y
A_a	Semimagor axis of an ellipse
A_b	Semiminor axis of an ellipse
B	Number of bootstrap samples
c	The parameter of the HQC
D	Total number of subjects in the three corneal groups
D_o	Distance between the object to the corneal vertex plan
D_{SE}	Domain of the structure element
$F\left(x, y\right)$	Clean unknown image
G_1	The MSE part of the score function for the information theoretic criteria
G_2	The penalty part of the score function for the information theoretic criteria
$\mathbf{G}_P$	The polar grid of the VKS image
$\mathbf{G}_C$	The Cartesian grid of the VKS image
$\mathbf{G}_{UN}$	A uniformally-spaced Cartesian grid of the VKS image
$h(x, y)$	2D impulse response
H_0	Null hypothesis
H_1	Alternative hypothesis
$H\left(\omega_1, \omega_2\right)$	2D frequency response of the Wiener filter
$I(x, y)$	VKS image
$I_N(x, y)$	Normalized VKS image
$I_P\left(x, y\right)$	The polar representation of the VKS image
$I_{PF}\left(x, y\right)$	The smoothed polar representation of the VKS image
$J_S(i, j)$	Squared sub-image around the central area of the VKS image
$J_E(i, j)$	The histogram-equalized version of $J_S(i, j)$
$J_{BW}(i, j)$	The edge contours of $J_E(i, j)$
$J_F(i, j)$	The region-filled version of $J_{BW}(i, j)$
$J_O(i, j)$	The image that contains the central object from which the centroid is detected
$\Omega(x, y)$	2D Gaussian function

$\mathbf{I}$	Identity matrix
k	Model-order
k_ψ	Number of parameters in the model ψ
K	Maximum allowed model-order
m	A scaling factor for the residuals of the bootstrap
m_F	Mean of the clean image
v_F	Variance of the clean image
m_N	Mean of the noise image
v_N	Variance of the noise image
n_a	Refractive index of the anterior corneal surface
n_ρ	Number of samples in the radial direction
n_θ	Number of samples in the angular direction
N	Number of samples
$N\left(x, y\right)$	Noise image
O_i	Size of the object's image
O_o	Size of the object
P_a	Refractive power of the anterior corneal surface
$P\left(\omega_1, \omega_2\right)$	2D power spectrum
$p_U(\boldsymbol{y})$	Universal distribution
$\boldsymbol{r}$	Vector of residuals
R_a	Anterior corneal radius of curvature
R_p	Posterior corneal radius of curvature
R_j	The tied ranks in corneal group j
$\boldsymbol{S}$	Vector of corneal height-data
$SE(i, j)$	Structure element for morphological operations
t_H	The higher value of the hysteresis thresholding
t_L	The lower value of the hysteresis thresholding
T_{KW}	Kruskal-Wallis test statistic
T_W	Wilcoxon test statistic
X_φ	Filled region
Z_p^q	Zernike polynomial with radial degree p and azimuthal frequency q
α	Nominal significance level for hypothesis testing
β	The parameter in the Mallows' $\mathrm{C_p}$ criterion
δ_ρ	The radial step in polar/Cartesian transformations for VKS images
δ_θ	The angular step in polar/Cartesian transformations for VKS images

$\delta\left(x,y\right)$	2D Dirac's delta function
$\varepsilon\left(\rho,\theta\right)$	Measurement and modeling error in the corneal height/elevation data
σ^2	Variance
$\hat{\sigma}_k^2$	Maximum-likelihood estimator of the variance of the surface residuals
γ	The parameter of the EDC
$\Gamma(\cdot)$	Euler's Gamma function
$(\cdot)^{\mathrm{T}}$	Transpose of a vector or matrix
$(\cdot)^{-1}$	Inverse of a square matrix
$\lvert \cdot \rvert$	Absolute value of a scalar
$\lvert\lvert \cdot \rvert\rvert$	Euclidean norm or 2-norm of a vector
$\ominus$	Morphological erosion
$\oplus$	Morphological dilation
$\circledast$	Hit-or-miss transformation
$\otimes$	Morphological thinning
$\bullet$	Morphological closing

Bibliography

[1] B. Cassin and S. Solomon, *Dictionary of Eye Terminology.* Gainsville, Florida: Triad Publishing Company, 1990.

[2] E. B. Goldstein, *Sensation & Perception*, 7th ed. Canada: Thompson Wadsworth, 2007.

[3] C. Roberts, “The accuracy of power maps to display curvature data in corneal topography systems,” *Invest. Ophthalmol. Vis. Sci.*, vol. 35, pp. 3525–3532, 1994.

[4] J. Schwiegerling, J. E. Greivenkamp, and J. M. Miller, “Representation of videokeratoscopic height data with Zernike polynomials,” *Journal of the Optical Society of America, A*, vol. 12, no. 10, pp. 2105–2113, October 1995.

[5] M. K. Smolek and S. D. Klyce, “Zernike polynomial fitting fails to represent all visually significant corneal aberrations,” *Investigative Ophthalmology and Visual Science*, vol. 44, pp. 4676–4681, 2003.

[6] D. R. Iskander, M. J. Collins, and B. Davis, “Optimal modeling of corneal surfaces with Zernike polynomials,” *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 87–95, January 2001.

[7] D. R. Iskander, M. J. Collins, B. Davis, and R. Franklin, “Corneal surface characterization: how many Zernike terms should be used?” *Investigat. Ophthalmol. Vis. Sci.*, vol. 42, no. 4, p. S896, 2001.

[8] D. R. Iskander, M. Morelande, M. J. Collins, and T. Buehren, “A refined bootstrap method for estimating the Zernike polynomial model order for corneal surfaces,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2203–2206, December 2004.

[9] A. Martinez-Finkelshtein, A. M. Delgado, G. M. Castro, A. Zarzo, and J. L. Alió, “Comparative analysis of some modal reconstruction methods of the shape of the cornea from corneal elevation data,” *Investigative Ophthalmology and Visual Science*, vol. 50, no. 12, pp. 5639–5645, December 2009.

[10] A. M. Zoubir, “Bootstrap methods for model selection,” *AEÜ International Journal of Electronics and Communications*, vol. 53, no. 6, pp. 386–392, December 1999.

[11] ——, “Model selection: a bootstrap approach,” in *The Proceedings of ICASSP’99*, Phoenix, Arizona, USA, March 1999, pp. 1377–1380.

[12] A. Zoubir and D. Iskander, *Bootstrap Techniques for Signal Processing.* Cambridge University Press, 2004.

[13] A. M. Zoubir and D. R. Iskander, “The bootstrap: A tutorial for the signal processing practitioner,” *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 10–19, July 2007.

[14] I. Vila, G. Cazuguel, A. L. Mehaute, F. Malet, and J. Colin, "A robust method for extracting and labelling mires in keratographic images," in *The Proceedings of the IEEE 17th Annual Conference of Engineering in Medicine and Biology Society*, vol. 2, 1995, pp. 1431–1432.

[15] J. M. Boi, R. Bulot, and J. Sequeira, "A new methodology to analyze images from videokeratography," in *The Proceedings of the IEEE 17th Annual Conference on Engineering in Medicine and Biology Society*, vol. 1, 1995, pp. 421–422.

[16] W. Alkhaldi, D. R. Iskander, A. M. Zoubir, and M. J. Collins, "Enhancing the standard operating range of a placido disk videokeratoscope for corneal surface estimation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 800–809, March 2009.

[17] W. Alkhaldi, D. R. Iskander, and A. M. Zoubir, "Model-order selection in Zernike polynomial expansion of corneal surfaces using the efficient detection criterion," *IEEE Transactions on Biomedical Engineering*, vol. xx, no. x, pp. xxxx–xxxx, xx 2010.

[18] ——, "Improving the performance of model-order selection criteria by partial-model selection search," in *The Proceedings of The 35th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Texas, USA, March 2010, pp. 4130–4133.

[19] ——, "The applicability of biased estimation in model and model order selection," in *The Proceedings of The 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009, pp. 3461–3464.

[20] D. R. Iskander, W. Alkhaldi, and A. M. Zoubir, "On the computer intensive methods in model selection," in *The Proceedings of The 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, March 2008, pp. 3461–3464.

[21] D. R. Iskander and W. Alkhaldi, "The "hook and loop" resampling plane," in *The Proceedings of The 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Virgin Islands, USA, December 2007, pp. 65–68.

[22] G. Smith and D. Atchison, *The eye and visual optical instruments.* Cambridge University Press, 1997.

[23] Y. L. Grand and S. G. E. Hage, *Physiological Optics.* Springer, 1980.

[24] C. W. Oyster, *The Human Eye: Structure and Function.* Sinauer Associates Inc. Publishers, 1999.

[25] D. A. Atchinson and G. Smith, *Optics of the human eye.* Butterworth-Heinemann, 2000.

[26] P. M. Kiely, G. Smith, and L. G. Carney, "The mean shape of the human cornea," *Optica Acta*, vol. 29, no. 8, pp. 1027–1040, August 1982.

[27] M. P. Keating, *Geometric, Physical and Visual Optics.* Butterworths-Heinemann, Stoneham, MA, 1988.

[28] S. A. Read, T. Buehren, and M. J. Collins, "Influence of accommodation on the anterior and posterior cornea," *Journal Cataract Refract. Surg.*, vol. 33, pp. 1877–1885, 2007.

[29] J. Németh, B. Erdélyi, and B. Csákány, "Corneal topography changes after a 15 second pause in blinking," *J. Cataract Refract Surg.*, vol. 27, no. 4, pp. 589–592, April 2001.

[30] D. J. Schanzlin and J. B. Robin, *Corneal Topography: Measuring and Modifying the Cornea.* Springer, 1992.

[31] J. F. Bille, C. F. H. Harner, and F. H. Loesel, *Aberration-Free Refractive Surgery: New Frontiers in Vision.* Springer, 2003.

[32] S. D. Klyce, "Computer-assisted corneal topography. high-resolution graphic presentation and analysis of keratoscopy," *Invest. Ophthamol. Vis. Sci.*, vol. 25, pp. 1426–1436, 1984.

[33] J. R. Levene, "The true inventors of the keratoscope and photokeratoscope," *The British Journal for the History of Science*, vol. 2, no. 8, pp. 324–341, 1965.

[34] P. Binder, "The true inventors of the keratoscope and photokeratoscope," *Journal of the Contact Lens Association of Ophthalmologists (CLAO)*, vol. 21, pp. 133–144, 1995.

[35] F. H. M. Jongsma, J. de Brabander, and F. Hendrikse, "Review and classification of corneal topographers," *Lasers in Medical Sciences*, vol. 14, no. 1, pp. 2–19, 1999.

[36] H. Helmholz, *Handbuch des physiologischen Optik.* Hamburg, Germany, Leopold Voss, 1909.

[37] J. W. Warnicki, "Corneal topography using computer analyzed rasterstereographic images," *Applied Optics*, vol. 27, pp. 1135–1140, 1988.

[38] M. W. Belin, D. Litoff, S. J. Strods, S. S. Winn, and R. S. Smith, "The par technology corneal topography system," *Refrect. Corneal. Surg.*, vol. 8, pp. 88–96, 1992.

[39] F. A. Guarnieri and J. C. Guarnieri, "Comparison of Placido-based, rasterstereography, and slit-scan corneal topography systems," *J. Refract. Surg.*, vol. 18, no. 2, pp. 169–176, 2002.

[40] M. W. Belin, D. Litoff, S. J. Strods, S. S. Winn, and R. S. Smith, "History and principles of Shack-Hartmann wavefront sensing," *Journal of Refractive Surgery*, vol. 17, pp. 573–577, 2001.

[41] S. A. Klein and R. B. Mandell, "Axial and instantaneous power conversion in corneal topography," *Investigative Ophthalmology and Visual Science*, vol. 36, pp. 2155–2159, 1995.

[42] ——, “Shape and refractive powers in corneal topography,” *Investigative Ophthalmology and Visual Science*, vol. 36, pp. 2096–2109, 1995.

[43] C. Roberts, “A practical guide to the interpretation of corneal topography,” *Contact Lens Spectrum*, vol. 3, pp. 25–33, 1998.

[44] S. A. Dingeldein and S. D. Klyce, “The topography of normal corneas,” *J. Cataract Refract. Surg.*, vol. 107, pp. 512–518, 1989.

[45] L. A. V. de Carvalho, A. C. R. ao, S. Tonissi, F. Yasuoka, J. C. Castro, P. Schor, and W. Chamon, “Videokeratograph (VKS) for monitoring corneal curvature during surgery,” *Arq. Bras. Oftalmo.*, vol. 65, pp. 37–41, 2002.

[46] P. R. Keller and P. P. van Saarloos, “Prespectives on corneal topography: a review on videokeratoscopy,” *Clinical and Experimental Optometry*, vol. 80, no. 1, pp. 18–33, 1997.

[47] J. D. Doss, R. L. Hutson, J. J. Rowsey, and D. R. Brown, “Method for calculation of corneal profile and power distribution,” *Arch. Ophthalmol.*, vol. 99, pp. 1261–1265, 1981.

[48] J. Wang, D. A. Rice, and S. D. Klyce, “A new reconstruction algorithm for improvement of corneal topographical analysis,” *J. Refract. Corneal Surg.*, vol. 5, pp. 379–387, 1989.

[49] S. A. Klein, “A corneal topography algorithm that produces continuous curvature,” *Optm. Vis. Sci.*, vol. 69, pp. 829–834, 1992.

[50] R. Mattioli and N. K. Tripoli, “Corneal geometry reconstruction with the keratron videokeratographer,” *Lasers in Medical Sciences*, vol. 74, no. 11, pp. 881–894, 1997.

[51] N. K. Tripoli, K. L. Cohen, D. E. Holmgren, and J. M. Coggins, “Assessment of radial aspheres by the arc-step algorithm as implemented by the keratron keratoscope,” *American Journal of Ophthalmology*, vol. 120, no. 5, pp. 658–654, 1995.

[52] J. Schwiegerling and J. E. Greivenkamp, “Using corneal height maps and polynomial decomposition to determine corneal aberrations,” *Optom. Vis. Sci.*, vol. 74, pp. 906–916, 1997.

[53] W. Tang, M. J. Collins, L. Carney, and B. Davis, “The accuracy and precision performance of four videokeratoscopes in measuring test surfaces,” *Optometry and Vision Science*, vol. 77, no. 9, pp. 483–491, 2000.

[54] D. R. Iskander and M. J. Collins, “Applications of high-speed videokeratoscopy,” *Clin. Exp. Optom.*, vol. 88, no. 4, pp. 223–231, 2005.

[55] J. Nemeth, B. Erdelyi, P. G. B. Csakany, A. Soumelidis, F. Kahlesh, and Z. Lang, “High-speed videotopographic measurement of tear film build-up time,” *Investigative Ophthalmology and Visual Science*, vol. 43, no. 6, pp. 1783–1790, 2002.

[56] D. R. Iskander and H. T. Kasprzak, "Dynamics in longitudinal eye movements and corneal shape," *Ophthal. Physiol. Opt.*, vol. 26, pp. 572–579, 2006.

[57] M. R. Morelande, D. R. Iskander, M. J. Collins, and R. Franklin, "Automatic estimation of the corneal limbus in videokeratoscopy," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 12, pp. 1617–1625, 2002.

[58] D. Alonso-Caneiro, D. R. Iskander, and M. J. Collins, "Estimating corneal surface topography in videokeratoscopy in the presence of strong signal interference," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 10, pp. 2381–2387, 2008.

[59] B. Seitz, A. Behrens, and A. L. A., "Corneal topography," *Curr. Opin. Ophthalmol.*, vol. 8, pp. 8–24, 1997.

[60] K. Marsolo, M. Twa, M. A. Bullimore, and S. Parthasarathy, "Spatial modeling and classification of corneal shape," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, pp. 203–212, 2007.

[61] J. Schwiegerling and J. E. G. JE, "Keratoconus detection based on videokeratoscopic height data," *Optom. Vis. Sci.*, vol. 73, pp. 721–728, 1996.

[62] M. Schneider, D. R. Iskander, and M. J. Collins, "Modeling corneal surfaces with rational functions for high-speed videokeratoscopy data compression," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 493–499, 2009.

[63] W. Lotmar, "Theoretical eye model with aspherics," *Journal of the Optical Society of America*, vol. 61, no. 11, pp. 1522–1529, 1971.

[64] H. T. Kasprzak and E. Jankowska-Kuchta, "A new analytical approximation of corneal topography," *Journal of Modern Optics*, vol. 43, no. 6, pp. 1135–1148, 1996.

[65] H. T. Kasprzak and D. R. Iskander, "Approximating ocular surfaces by generalised conic curves," *Ophthalmic and Physiological Optics*, vol. 26, no. 6, pp. 602–609, 2006.

[66] H. C. Howland, A. Glasser, and R. Applegate, "Polynomial approximations of corneal surfaces and corneal curvature topography," in *Ophthalmic and Visual Optics of the 1992 OSA Technical Digest Series, Optical Society of America*, vol. 3, 1992, pp. 34–37.

[67] D. R. Iskander, M. R. Morelande, M. J. Collins, and B. Davis, "Modeling corneal surfaces with radial polynomials," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 4, pp. 320–328, 2002.

[68] D. R. Iskander, "Modeling videokeratoscopic height data with spherical harmonics," *Optometry and Vision Science*, vol. 86, no. 5, pp. 542–547, 2009.

[69] R. H. Webb, "Zernike polynomial description of ophthalmic surfaces," in *Ophthalmic and Visual Optics of the 1992 OSA Technical Digest Series, Optical Society of America*, vol. 3, 1992, pp. 38–41.

[70] C.-J. Kim and R. R. Shannon, *Catalog of Zernike polynomials.* Academic, New York: Applied Optics and Optical Engineering, R. R. Shannon and J. Wyant, eds., 1992.

[71] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[72] *Methods for Reporting Optical Aberrations of Eyes*, American National Standards Institute, ANSI Z80.28–2004.

[73] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Amer.*, vol. 66, no. 3, pp. 207–211, 1976.

[74] C. Roberts, "Characterization of inherent error in a spherically-biased corneal topography system in mapping a radially aspheric surface," *Journal of Refractive and Corneal Surgery*, vol. 10, no. 2, pp. 103–116, 1994.

[75] S. A. Klein, "A corneal topography reconstruction algorithm that avoids the skew ray ambiguity and the skew ray error," *Optometry and Vision Science*, vol. 74, no. 11, pp. 945–962, 1997.

[76] D. R. Iskander, B. Davis, and M. J. Collins, "The skew ray ambiguity in the analysis of videokeratoscopic data," *Optometry and Vision Science*, vol. 84, no. 5, pp. 435–442, 2007.

[77] R. A. Applegate, R. Nunez, J. Buettner, and H. C. Howland, "How accurately can videokeratographic system measure surface elevation," *Optometry and Vision Science*, vol. 72, no. 11, pp. 785–792, 1995.

[78] P. Cho, A. K. C. Lam, J. Mountford, and L. Ng, "The performance of four different corneal topographers on normal human corneas and its impact on orthokeratology lens fitting," *Optometry and Vision Science*, vol. 79, no. 3, pp. 175–183, 2002.

[79] T. T. McMahon, R. J. Anderson, C. E. Joslin, and G. A. Rosas, "Precision of three topography instruments in keratoconus subjects," *Optometry and Vision Science*, vol. 78, no. 8, pp. 599–604, 2001.

[80] J. Andersen, P. Koch-Jensen, and O. Osterby, "Corneal topography: Image processing and numerical analysis of keratoscopy," *Acta Ophthalmologica*, vol. 71, no. 2, pp. 151–159, 1993.

[81] L. Wang and G. U. Auffarth, "White-to-white corneal diameter measurements using the eyemetrics program of the orbscan topography system," *Developments in Ophthalmology*, vol. 34, pp. 141–146, 2002.

[82] D. R. Iskander, M. J. Collins, and B. Davis, "Evaluating tear film stability in the human eye with high-speed videokeratoscopy," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1939–1949, 2005.

[83] K. Tsubota, "Tear dynamics and dry eye," *Progress in Retinal and Eye Research*, vol. 17, no. 4, pp. 556–596, 1998.

[84] R. E. Woods and R. C. Gonzalez, “Real-time digital image enhancement,” *Proc. IEEE*, vol. 69, no. 5, pp. 643–654, May 1981.

[85] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, November 1986.

[86] L. Vincent, “Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms,” *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 176–201, April 1993.

[87] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Prentice-Hall, Inc., 2002.

[88] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision.* Addision-Wesley, 1992, vol. 1.

[89] L. Lam, S. W. Lee, and C. Y. Suen, “Thinning methodologies–a comprehensive survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869–885, September 1992.

[90] E. Kreyszig, *Advanced Engineering Mathematics*, 7th ed. New York: Wiley, 1993.

[91] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* Cambridge, Mass: MIT Press, 1964.

[92] J. S. Lim, *Two-Dimensional Signal and Image Processing.* Prentice-Hall, Inc., 1990.

[93] B. Delaunay, “Sur la sphère vide,” *Otdelenie Matematicheskikh i Estestvennykh Nauk*, vol. 7, pp. 793–800, 1934.

[94] G. Rawlins and D. Wood, “Ortho-convexity and its generalizations,” *Computational Morphology*, pp. 137–152, 1988.

[95] C. Barber, D. P. Dobkin, and H. T. Huhdanpaa, “The quickhull algorithm for convex hulls,” *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, December 1996.

[96] D. Watson, *Contouring: A Guide to the Analysis and Display of Spatial Data.* Pergamon, 1994.

[97] J. Shao and D. Tu, *The jackknife and bootstrap.* Springer Verlag, New York, 1995.

[98] D. N. Politis, J. P. Romano, and M. Wolf, *Subsampling.* Springer, 1999.

[99] A. Guirao and P. Artal, “Corneal wave aberration from videokeratography: accuracy and limitations of the procedure,” *Journal of the Optical Society of America, A*, vol. 17, no. 6, pp. 955–965, June 2000.

[100] N. Maeda, S. D. Klyce, and M. K. Smolek, "Comparison of methods for detecting keratoconus using videokeratography," *Archives of Ophthalmology*, vol. 113, no. 7, pp. 870–874, July 1995.

[101] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[102] C. M. Hurvich and C.-L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.

[103] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[104] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *J. R. Statist. Soc. B*, vol. 41, no. 2, pp. 190–195, 1979.

[105] J. Rissanen, "MDL denoising," *IEEE Transactions On Information Technology*, vol. 46, no. 7, pp. 2537–2543, 2000.

[106] S. Kay, "Conditional model order estimation," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1910–1917, 2001.

[107] C. L. Mallows, "Some comments on Cp," *Technometrics*, vol. 15, pp. 661–675, 1973.

[108] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, pp. 1–25, 1986.

[109] C. R. Rao and Y. Wu, "A strongly consistent procedure for model selection in a regression problem," *Biometrika*, vol. 76, no. 2, pp. 369–374, 1989.

[110] J. Shao, "Bootstrap model selection," *Journal of American Statistical Association*, vol. 91, no. 434, pp. 655–665, June 1996.

[111] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Statist. Ass.*, vol. 47, pp. 583–621, 1952.

[112] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.

[113] S. M. Kendall and A. Stuart, *The Advanced Theory of Statistics.* New York: Macmillan, 1977, vol. 2.

[114] K. O. Dzhaparidze and A. M. Yaglom, "Spectrum parameter estimation in time series analysis," *Developments in statistics*, 1983.

[115] L. Ljung, *System identification: theory for the user.* Prentice-Hall, Englewood Cliffs, NJ, 1987.

[116] A. W. Rihaczek, *Principles of high-resolution radar.* Peninsula Publishing, Los Atlos, CA, 1985.

[117] A. M. Zoubir and D. R. Iskander, “Bootstrap modeling of a class of nonstationary signals,” *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 399–408, 2000.

[118] H. White and J. Racine, “statistical inference, the bootstrap, and neural-network modeling with applications to foreign exchange rates,” *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 657–673, 2001.

[119] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory.* New York: Wiley, 1994.

[120] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1998.

[121] S. Kay and Y. Eldar, “Rethinking biased estimation,” *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, 2008.

[122] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[123] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference.* Springer, 1998.

[124] J. Myung, D. Navarro, and M. Pitt, “Model selection by normalized maximum likelihood,” *Journal of Mathematical Psychology*, vol. 50, pp. 167–179, 2006.

[125] J. Rissanen, “Fisher information and stochastic complexity,” *IEEE Transactions On Information Technology*, vol. 42, pp. 40–47, January 1996.

[126] G. Claeskens and N. L. Hjort, *Model Selection And Model Averaging.* Cambridge University Press, 2008.

[127] P. M. T. Broersen, “How to select polynomial models with an accurate derivative,” *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 5, pp. 910–914, October 2000.

[128] P. M. T. Broersen and H. E. Wensink, “On the penalty factor for autoregressive order selection in finite samples,” *IEEE Transactions On Signal Processing*, vol. 44, pp. 748–752, 1996.

[129] W. Chong-Yaw, R. Paramesran, and F. Takeda, “Bootstrap model order selection of Zernike polynomial expansion for classification of rice,” in *The Proceedings of TENCON’04*, vol. 1, 2004, pp. 203–206.

[130] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap.* Chapman and Hall, New York, 1993.

[131] B. Carlson, “Covariance matrix estimation errors and diagonal loading in adaptive arrays,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 3.

[132] B. J. N. Blight, “Some general results on reduced mean square error estimation,” *American Statistician*, vol. 25, no. 3.

[133] E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 42, no. 1, pp. 80–86, Feb. 2000.

[134] P. C. Hansen, "Regularization tools: a matlab package for analysis and solution of discrete ill-posed problems," *Numerical Algorithms*, vol. 6, pp. 1–35, 1994.

[135] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc B.*, vol. 58, no. 1, pp. 267–288, 1996.

[136] Y. C. Eldar, "Uniformaly improving the Cramer-Rao bound and the maximum likelihood estimation," *IEEE Transactions On Signal Processing*, vol. 54, pp. 2943–2956, August 2006.

[137] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean squared estimation in the presence of model uncertainties," *IEEE Transactions On Signal Processing*, vol. 53, pp. 168–181, January 2005.

Lebenslauf

Name: Weaam F. H. Alkhaldi
Geburtsdatum: 10 Juni 1974
Geburtsort: Alexandria, Ägypten
Staatsangehrigkeit: Palästinenser
Familienstand: Verheiratet

Studium

05/1998-12/2002	Studium der Elektrotechnik an der Arab Academy for Science and Technology, Ägypten. Studienabschluß: Master Degree of Electronics and Communication Engineering.
03/1993-07/1997	Studium der Elektrotechnik an der University of Garyounis, Libyen. Studienabschluß: B.Sc. Degree in Electrical and Electronics Engineering.
07/1992	Abschluss des Gymnasiums (Abitur), Ägypten.

Berufstätigkeit

03/2010 -	Wissenschaftlicher Mitarbeiter am Fachgebiet Signalverarbeitung, Institut für Nachrichtentechnik, Technische Universität Darmstadt.
05/2006 - 02/2010	DAAD Stipendiat und Wissenschaftlicher Mitarbeiter am Fachgebiet Signalverarbeitung, Institut für Nachrichtentechnik, Technische Universität Darmstadt.
01/2006-04/2006	DAAD Stipendiat und Wissenschaftlicher Mitarbeiter am Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg.
09/2003-06/2005	Dozent und Wissenschaftlicher Mitarbeiter am Fachgebiet Elektrotechnik, Islamic University-Gaza, Palästina.
02/2002-08/2003	Tutor am Fachgebiet Elektrotechnik, Arab Academy for Science and Technology, Ägypten.